Institute of Mathematical Statistics
LECTURE NOTES–MONOGRAPH SERIES

Weighted Empiricals and Linear Models

Hira L. Koul
Michigan State University

Volume 21

Institute of Mathematical Statistics
LECTURE NOTES–MONOGRAPH SERIES
Volume 21

Weighted Empiricals and Linear Models

Hira L. Koul
Michigan State University

Institute of Mathematical Statistics
Hayward, California

Institute of Mathematical Statistics

Lecture Notes–Monograph Series

The production of the *IMS Lecture Notes–Monograph Series* is managed by the IMS Business Office: Jessica Utts, IMS Treasurer, and Jose L. Gonzalez, IMS Business Manager.

Library of Congress Catalog Card Number: 92-54658

International Standard Book Number 0-940600-28-5

Printed in the United States of America

To the memory of my parents

Prabhawati (Dhar) and Radhakrishen Koul (Gassi)

PREFACE

An empirical process that assigns possibly different non-random (random) weights to different observations is called a *weighted* (*randomly weighted*) *empirical process*. These processes are as basic to linear regression and autoregression models as the ordinary empirical process is to one sample models. However their usefulness in studying linear regression and autoregression models has not been fully exploited. This monograph addresses this question to a large extent.

There is a vast literature in Nonparametric Inference that discusses inferential procedures based on empirical processes in k-sample location models. However, their analogs in autoregression and linear regression models are not readily accessible. This monograph makes an attempt to fill this void. The statistical methodologies studied here extend to these models many of the known results in k-sample location models, thereby giving a unified theory.

By viewing linear regression models via certain weighted empirical processes one is naturally led to new and interesting inferential procedures. Examples include minimum distance estimators of regression parameters and goodness-of-fit tests pertaining to the errors in linear models. Similarly, by viewing autoregression models via certain randomly weighted empirical processes one is naturally led to classes of minimum distance estimators of autoregression parameters and goodness-of-fit tests pertaining to the error distribution.

The introductory Chapter 1 gives an overview of the usefulness of weighted and randomly weighted empirical processes in linear models. Chapter 2 gives general sufficient conditions for the weak convergence of suitably standardized versions of these processes to continuous Gaussian processes. This chapter also contains the proof of the asymptotic uniform linearity of weighted empirical processes based on the residuals when errors are heteroscedastic and independent. Chapter 3 discusses the asymptotic uniform linearity of linear rank and signed rank statistics when errors are heteroscedastic and independent. It also includes some results about the weak convergence of weighted empirical processes of ranks and signed ranks. Chapter 4 is devoted to the study of the asymptotic behavior of M- and R-estimators of regression parameters under heteroscedastic and independent errors, via weighted empirical processes. A brief discussion about bootstrap approximations to the distribution of a class of M-estimators appears in Section 4.2b. This chapter also contains a proof of the consistency of a class of robust estimators for certain scale parameters under heteroscedastic errors.

In carrying out the analysis of variance of linear regression models based on ranks, one often needs an estimator of the functional $\int f d\varphi(F)$, where F is the error distribution function, f its density and φ is a function from $[0, 1]$ to the real line. Some estimators of this functional and the proofs of their consistency in the linear regression setting appear in Section 4.5.

Chapters 5 and 6 deal with minimum distance estimation, via weighted empirical processes, of the regression parameters and tests of goodness-of-fit pertaining to the error distribution. One of the main themes emerging from these two chapters is that the inferential procedures based on weighted empiricals with weights proportional to the design matrix provide the right extensions of k-sample location model procedures to linear regression models.

It is customary to expect that a method that works for linear regression models should have an analogue that will also work in autoregression models. Indeed many of the inferential procedures based on weighted empirical processes in linear regression that are discussed in Chapters 3-6 have precise analogs in autoregression based on certain randomly weighted empirical processes and appear in Chapter 7. In particular, the proof of the asymptotic uniform linearity of the ordinary empirical process of the residuals in autoregression appears here.

All asymptotic uniform linearity results in the monograph are shown to be consequences of the asymptotic continuity of certain basic weighted and randomly weighted empirical processes.

Chapters 2-4 are interdependent. Chapter 5 is mostly self-contained and can be read after reading the Introduction. Chapter 6 uses results from Chapters 2 and 5. Chapter 7 is almost self-contained. The basic result needed for this chapter appears in Section 2.2b.

The first version of this monograph was prepared while I was visiting the Department of Statistics, Poona University, India, on sabbatical leave from Michigan State University, during the academic year 1982–83. Several lectures on some parts of this monograph were given at the Indian Statistical Institute, New Delhi, and Universities of La Trobe, Australia, and Wisconsin, Madison. I wish to thank Professors S. R. Adke, Richard Johnson, S. K. Mitra, M. S. Prasad and B. L. S. Prakasa Rao for having some discussions pertaining to the monograph. My special thanks go to James Hannan for encouraging me to finish the project and for proof reading parts of the manuscript, to Soumendra Lahiri for helping me with sections on bootstrapping, and to Bob Serfling for taking keen interest in the monograph and for many comments that helped to improve the initial draft.

Ms. Achala Sabane and Ms. Lora Kemler had the pedestrian task of typing the manuscript. Their patient endeavors are gratefully acknowledged. Ms. Kemler's keen eye for details has been an indispensable help.

During the preparation of the monograph the author was partly supported by the University Grants Commission of India and the National Science Foundation, grant numbers NSF 82–01291, DMS–9102041.

Hira L. Koul,
East Lansing, MI. 48823
5/28/92.

NOTATION AND CONVENTIONS

The p–dimension Euclidean space is denoted by $\mathbb{R}^p$, $p \geq 1$; $\mathbb{R} = \mathbb{R}^1$; $\mathscr{B}^p :=$ the σ–algebra of Borel sets in $\mathbb{R}^p$, $\mathscr{B} = \mathscr{B}^1$; $\lambda :=$ Lebesgue measure on $(\mathbb{R}, \mathscr{B})$. The symbol ":=" stands for "by definition".

For any set $A \subset \mathbb{R}$, $\mathbb{D}(A)$ denotes the class of real valued functions on A that are right continuous and have left limits while $\mathbb{D}\mathcal{I}(A)$ denotes the subclass in $\mathbb{D}(A)$ whose members are nondecreasing. $\mathbb{C}[0, 1] :=$ the class of real valued bounded continuous functions on $[0, 1]$.

A vector or a matrix will be designated by a bold letter. A $\mathbf{t} \in \mathbb{R}^p$ is a p×1 vector, $\mathbf{t}'$ or $\mathbf{t}^{\prime}$ its transpose, $\|\mathbf{t}\|^2 := \sum_{j=1}^{p} t_j^2$, $|\mathbf{t}| := \max\{|t_j|, 1 \leq j \leq p\}$. For any p–square matrix $\mathbf{C}$, $\|\mathbf{C}\|_\infty = \sup\{\|\mathbf{t}'\mathbf{C}\|;\ \|\mathbf{t}\| \leq 1\}$. For an n×p matrix $\mathbf{D}$, $\mathbf{d}_{ni}^{\prime}$ denotes its i^{th} row, $1 \leq i \leq n$, and $\mathbf{D}_c$ the n×p matrix $\mathbf{D} - \overline{\mathbf{D}}$, whose i^{th} row consists of $(\mathbf{d}_{ni} - \bar{\mathbf{d}}_n)'$, with $\bar{\mathbf{d}}_n := \Sigma_i \mathbf{d}_{ni}/n$, $1 \leq i \leq n$.

w.e.p.('s)	:= weighted empirical process(es).
r.w.e.p.('s)	:= randomly weighted empirical process(es).
i.i.d.	:= independent identically distributed.
r.v.('s)	:= random variable(s).
d.f.('s)	:= distribution function(s).
w.r.t.	:= with respect to.
C-S	:= the Cauchy–Schwarz inequality.
D.C.T.	:= the Dominated Convergence Theorem
Fubini	:= the Fubini Theorem.
L-F CLT	:= the Lindeberg–Feller Central Limit Theorem.
$o(1)(o_p(1))$	:= a sequence of numbers (r.v.'s) converging to zero (in probability).
$O(1)(O_p(1))$	:= a sequence of numbers (r.v.'s) that is bounded (in probability).
$N(\mathbf{0}, \mathbf{C})$	:= either a r.v. with normal distribution whose mean vector is $\mathbf{0}$ and the covariance matrix $\mathbf{C}$ or the corresponding distribution.
$\|g\|_\infty$	:= the supremum norm over the domain of g, g a real valued function.
τ_a^2	:= $\sum_{i=1}^{n} a_{ni}^2$, for an arbitrary real vector $(a_{n1}, \ldots, a_{nn})'$.

Often in a discussion or in a proof the subscript n on the triangular arrays and various other quantities will not be exhibited. The index i in Σ_i or $\sum_i$ and $\max_i$ or $\max_i$ will vary from 1 to n, unless specified otherwise. All limits, unless specified otherwise, are taken as $n \longrightarrow \infty$.

For a sequence of r.v.'s $\{X, X_n, n \ge 1\}$, $X_n \underset{d}{\to} X$ means that the distribution of X_n converges weakly to that of X. For two r.v.'s X, Y, $X \underset{d}{=} Y$ means that the distribution of X is the same as that of Y.

For a sequence of stochastic processes $\{Y, Y_n, n \ge 1\}$, $Y_n \Rightarrow Y$ means that Y_n converges weakly to Y in a given topology. $Y_n \underset{fd}{\to} Y$ means that all finite dimensional distributions of Y_n converge weakly to that of Y.

Reference to an expression or a display is made by the (expression number) if referring in the same section and by the (chapter number.section number.expression number), otherwise. For example, by (3.2.1) is meant an expression (1) of Section 2 of Chapter 3. A reference to this while in Section 3.2 would appear as (1).

For convenient reference we list here some of the most often used conditions in the manuscript. For an arbitrary d.f. F on $\mathbb{R}$, conditions **(F1)**, **(F2)** and **(F3)** are as follows:

(F1) F has uniformly continuous density f w.r.t. λ.

(F2) $f > 0$, a.e. λ.

(F3) $\sup_{x \in \mathbb{R}} |xf(x)| < \infty$.

These conditions are introduced for the first time just before Corollary 3.2.1 and are used frequently subsequently.

For an $n \times p$ design matrix matrix $\mathbf{X}$, the conditions (NX), (NX1) and (NX$_c$) are as follows:

(NX) $$(\mathbf{X}'\mathbf{X})^{-1} \text{ exists}, \ n \ge p; \quad \max_i \mathbf{x}_{ni}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_{ni}' = o(1).$$

(NX1) $$(\mathbf{X}_c'\mathbf{X}_c)^{-1} \text{ exists}, \ n \ge p;$$
$$\max_i \mathbf{x}_{ni}'(\mathbf{X}_c'\mathbf{X}_c)^{-1}\mathbf{x}_{ni} = o(1).$$

(NX$_c$) $$(\mathbf{X}_c'\mathbf{X}_c)^{-1} \text{ exists}, \ n \ge p;$$
$$\max_i (\mathbf{x}_{ni} - \bar{\mathbf{x}}_n)'(\mathbf{X}_c'\mathbf{X}_c)^{-1}(\mathbf{x}_{ni} - \bar{\mathbf{x}}_n) = o(1).$$

The condition (NX) is the most often used from Theorem 2.3.3. onwards. The letter N in these conditions stands for Noether, who was the first person to use (NX), in the case p=1, to obtain the asymptotic normality of weighted sums of r.v.'s; see Noether (1949).

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

1.1. WEIGHTED EMPIRICAL PROCESSES

A ***weighted empirical process*** (w.e.p.) corresponding to the random variables (r.v.'s) $X_{n1}, \ldots, X_{nn}$ and the non–random real weights $d_{n1}, \ldots, d_{nn}$ is defined to be

$$U_d(x) := \sum_{i=1}^{n} d_{ni}\, I(X_{ni} \le x), \qquad x \in \mathbb{R},\ n \ge 1.$$

The weights $\{d_{ni}\}$ need not be nonnegative.

The classical example of a w.e.p. is the *ordinary empirical process* that corresponds to $d_{ni} \equiv n^{-1}$. Another example is given by the two sample empirical process obtained as follows: Let m be an integer, $1 \le m \le n$, $r := n - m$; $d_{ni} = -r/n$, $1 \le i \le m$; $d_{ni} = m/n$, $m + 1 \le i \le n$. Then the corresponding U_d-process becomes

$$U_d(x) \equiv (mr/n)\, \{r^{-1} \sum_{i=m+1}^{n} I(X_{ni} \le x) - m^{-1} \sum_{i=1}^{m} I(X_{ni} \le x)\}, \qquad x \in \mathbb{R},$$

precisely the process that arises in two-sample models.

More generally, weighted empirical processes (w.e.p.'s) arise naturally in linear regression models where, for each $n \ge 1$ and each $\boldsymbol{\beta} \in \mathbb{R}^p$, the data $\{(\mathbf{x}_{ni}', Y_{ni}), 1 \le i \le n\}$ are related to the error variables $\{e_{ni}, 1 \le i \le n\}$ by the linear relation

$$Y_{ni} = \mathbf{x}_{ni}'\boldsymbol{\beta} + e_{ni}, \quad 1 \le i \le n. \tag{1}$$

Here $e_{n1}, \ldots, e_{nn}$ are independent r.v.'s with respective continuous d.f.'s $F_{n1}, \ldots, F_{nn}$, $\mathbf{x}_{ni}' = (x_{ni1}, \ldots, x_{nip})$ is the i^{th} row of the known $n{\times}p$ design matrix $\mathbf{X}$ and $\boldsymbol{\beta}$ is the parameter vector of interest.

Consider the vector of w.e.p.'s $\mathbf{V} := (V_1, \ldots, V_p)'$ where

$$V_j(y, \mathbf{t}) := \sum_{i=1}^{n} x_{nij}\, I(Y_{ni} \le y + \mathbf{x}_{ni}'\mathbf{t}), \quad y \in \mathbb{R},\ \mathbf{t} \in \mathbb{R}^p,\ 1 \le j \le p. \tag{2}$$

Clearly, $V_j(\cdot, \mathbf{t})$ is an example of the w.e.p. $U_d(\cdot)$ with $d_{ni} \equiv x_{nij}$ and $X_{ni} \equiv Y_{ni} - \mathbf{x}_{ni}'\mathbf{t}$, $1 \le i \le n$, $1 \le j \le p$.

Observe that the data $\{(\mathbf{x}_{ni}', Y_{ni}), 1 \le i \le n\}$ in the model (1) are readily summarized by the vector of w.e.p.'s $\{\mathbf{V}(y, \mathbf{0}), y \in \mathbb{R}\}$ in the sense

that the given data can be recovered from the sample paths of this vector up to a permutation. This in turn suffices for the purpose of inference about β in (1). In this sense the vector of w.e.p.'s $\{V(y, 0), y \in \mathbb{R}\}$ is at least as important to linear regression models (1) as is the ordinary empirical process to one-sample location models. One of the purposes of this monograph is to discuss the role of V-processes in inference and in proving limit theorems in models (1) in a unified fashion.

1.2. M-, R- AND SCALE ESTIMATORS

Many inferential procedures involving (1.1.1) can be viewed as functions of V. For example the least squares estimator, or more generally, the class of M-estimators corresponding to the score function ψ, (Huber: 1981), is defined as a solution $\mathbf{t}$ of the equation

$$\int \psi(y)\, V(dy, \mathbf{t}) = \text{a known constant.}$$

Similarly, rank (R) estimators of β corresponding to the score function φ are defined to be a solution $\mathbf{t}$ of the equation

$$\int \varphi(H_n(y, \mathbf{t}))V(dy, \mathbf{t}) = \text{a known constant,} \tag{1}$$

$$H_n(y, \mathbf{t}) := n^{-1} \sum_{i=1}^{n} I(Y_{ni} \le y + \mathbf{x}_{ni}'\mathbf{t}), \qquad y \in \mathbb{R},\ \mathbf{t} \in \mathbb{R}^p.$$

A significant portion of Nonparametric Inference in models (1.1.1) deals with M- and R- estimators of β (Adichie; 1967. Huber; 1973) and linear rank tests of hypotheses about β, (Hájek–Sidák; 1967). By viewing these procedures as functions of $\{V(y, \mathbf{t}), y \in \mathbb{R}, \mathbf{t} \in \mathbb{R}^p\}$, it is possible to give a unified treatment of their asymptotic distribution theory, as is done in Chapters 3 and 4 below.

There is a vast literature in Nonparametric Inference that discusses inferential procedures based on functionals of empirical processes in the k-sample location model such as the books by Puri and Sen (1969), Serfling (1980) and Huber (1981). Yet their appropriate extensions to the linear regression model are not readily accessible. This monograph seeks to fill this void. The methodology and inference procedures studied here extend many known results in the k-sample location model to the model (1.1.1), thereby giving a unified treatment.

An important result needed for study of the asymptotic behavior of R-estimators of β is the asymptotic uniform linearity of the linear rank statistics of (1) in the regression parameter vector. Jurečková (1969, 1971) obtained this result under (1.1.1) with i.i.d. errors. A similar result was proved in Koul (1969, 1971) and Van Eeden (1972) for linear signed rank statistics under i.i.d. symmetric errors. Its extension to the case of

nonidentically distributed errors is not readily available. Theorems 3.2.4 and 3.3.3 prove the asymptotic uniform linearity of linear rank and linear signed rank statistics with bounded scores under the general independent errors model (1.1.1). In the case of i.i.d. errors, the conditions in these theorems on the error d.f. are more general than requiring finite Fisher information. The results are proved uniformly over all bounded score functions and are consequences solely of the asymptotic sample continuity of V-processes and some smoothness of $\{F_{ni}\}$. The uniformity with respect to the score functions is useful when constructing adaptive rank tests that are asymptotically efficient against Pitman alternatives for a large class of error distributions.

Chapter 3 also contains a proof of the asymptotic normality of linear rank and linear signed rank statistics under independent alternatives and for indicator score functions. This proof proceeds via the weak convergence of certain basic w.e.p.'s and complements some of the results in Dupač and Hájek (1969).

Section 4.2a discusses the asymptotic distribution of M-estimators under heteroscedastic errors using the asymptotic continuity of V-processes. Section 4.2b presents some second order results on bootstrap approximations to the distributions of a class of M-estimators.

In order to make M-estimators scale invariant one often needs an appropriate robust scale estimator. One such scale estimator, as recommended by Huber (1981) and others, is

$$s_1 = \operatorname{med}\{|Y_{ni} - \mathbf{x}_{ni}'\hat{\boldsymbol{\beta}}|, \quad 1 \le i \le n\},$$

where $\hat{\boldsymbol{\beta}}$ is an estimator of $\boldsymbol{\beta}$. The asymptotic distribution of s_1 under heteroscedastic errors is given in Section 4.3. In the case of i.i.d. errors, this asymptotic distribution does not depend on $\hat{\boldsymbol{\beta}}$ provided the errors are symmetric around 0. This observation naturally leads one to construct a scale estimator based on the symmetrized residuals, thereby giving another scale estimator

$$s_2 := \operatorname{med}\{|Y_{ni} - \mathbf{x}_{ni}'\hat{\boldsymbol{\beta}} - Y_{ni} + \mathbf{x}_{ni}'\hat{\boldsymbol{\beta}}|; \ 1 \le i, j \le n\}.$$

As expected, the asymptotic distribution of s_2 is shown to be free from the estimator $\hat{\boldsymbol{\beta}}$ in the case of i.i.d. errors, not necessarily symmetric. It also appears in Section 4.3.

Section 4.4 discusses the asymptotic distribution of a class of R-estimators under heteroscedastic errors using the asymptotic uniform linearity results of Chapter 3. The R-estimators considered are asymptotically equivalent to Jaeckel's estimators.

The complete rank analysis of the linear regression model (1.1.1) requires an estimate of the scale parameter

$$Q(f) := \int f \, d\varphi(F)$$

where f is density of the unknown common error d.f. F and φ is a nondecreasing function on (0, 1). This estimate is used to standardize the test statistic and estimate the standard error of the R-estimator corresponding to the score function φ. This parameter also appears in the efficiency comparisons of rank procedures and it is of interest to estimate it, after the fact, in an analysis.

Lehmann (1963), Sen (1966), Koul (1971), among others, provide estimators of Q(f) in the one- and two- sample location models and in the linear regression model. These estimators are given in terms of the lengths or Lebesgue measures of certain confidence intervals or regions. They are usually not easy to compute when the dimension p of β is larger than 1.

In Section 4.5, estimators of Q(f), based on kernel type density estimators of f and the empirical d.f. H_n, are defined and their consistency under (1.1.1) with i.i.d. errors is proved. An estimator whose window width is based on the data and is of the order of square root n, is also considered. The consistency proof presented is a sole consequence of the asymptotic continuity of certain w.e.p.'s and some smoothness of the error d.f.'s.

1.3. MINIMUM DISTANCE ESTIMATORS AND GOODNESS–OF–FIT TESTS

The practice of obtaining estimators of parameters by minimizing a certain distance between some functions of observations and parameters has been present in statistics since its beginning. The classical examples of this method are the Least Square and the minimum Chi Square estimators.

The minimum distance estimation (m.d.e.) method, where one obtains an estimator of a parameter by minimizing some distance between the empirical d. f. and the modeled d. f., was elevated to a general method of estimation by Wolfowitz (1953, 1954, 1957). In these papers he demonstrated that, compared to the maximum likelihood estimation method, the m.d.e. method yielded consistent estimators rather cheaply in several problems of varied levels of difficulty.

This methodology saw increasing research activity from the mid 1970's when many authors demonstrated various robustness properties of certain m.d. estimators. See, e.g., Beran (1977, 1978), Parr and Schucany (1979), Millar (1981, 1982, 1984), Donoho and Liu (1988 a, b), among others. All of these authors restrict their attention to the one sample setup or to the two sample location model. See Parr (1981) for additional bibliography on m.d.e. till 1980.

Inspite of many advances made in the m.d.e. methodology in one sample models, little was known till early 1980's as to how to extend this methodology to one of the most applied models, v.i.z., the multiple linear regression model (1.1.1). A ***significant*** advantage of viewing the model (1.1.1) through **V** is that one is naturally led to interesting m.d. estimators of β that are natural extensions of their one- and two- sample location model counterparts. To illustrate this, consider the m.d. estimator $\hat{\theta}$ of the

one sample location parameter θ, when errors are i.i.d. symmetric around 0, defined by the relation

$$\hat{\theta} := \operatorname{argmin}\{T_n(t);\ t \in \mathbb{R}\},$$

with

$$T_n(t) = \int \{n^{-1/2} \sum_{i=1}^{n} I(Y_{ni} \le y + t) - I(-Y_{ni} < y - t)]\}^2 dG(y), \quad t \in \mathbb{R},$$

where $G \in \mathbb{D}\mathcal{I}(\mathbb{R})$. Since (1.1.1) is an extension of the one sample location model, it is only natural to seek an extension of $\hat{\theta}$ in this model. Assuming that $\{e_{ni}\}$ are symmetrically distributed around 0, the first thing one is tempted to consider as an extension of $\hat{\theta}$ is β_1^+ defined by the relation

$$\beta_1^+ := \operatorname{argmin}\{K_1^+(\mathbf{t});\ \mathbf{t} \in \mathbb{R}^p\},$$

with

$$K_1^+(\mathbf{t}) = \int \{n^{-1/2} \sum_{i=1}^{n} I(Y_{ni} \le y + \mathbf{x}_{ni}'\mathbf{t}) - I(-Y_{ni} < y - \mathbf{x}_{ni}'\mathbf{t})]\}^2 dG(y), \quad \mathbf{t} \in \mathbb{R}^p.$$

However, any extension of $\hat{\theta}$ to the linear regression model should have the property that it reduce to $\hat{\theta}$ when the model is reduced to the one sample location model and, in addition, that it reduce to an appropriate extension of $\hat{\theta}$ to the k-sample location model when the model (1.1.1) is reduced to it. In this sense β_1^+ does not provide the right extension but $\beta_{\mathbf{X}}^+$ does, where

$$\text{(1)} \qquad \beta_{\mathbf{X}}^+ := \operatorname{argmin}\{K_{\mathbf{X}}^+(\mathbf{t});\ \mathbf{t} \in \mathbb{R}^p\},$$

with

$$K_{\mathbf{X}}^+(\mathbf{t}) := \int \mathbf{V}^{+\prime}(y, \mathbf{t})\, (\mathbf{X}'\mathbf{X})^{-1}\, \mathbf{V}^+(y, \mathbf{t})\, dG(y), \qquad \mathbf{t} \in \mathbb{R}^p,$$

$$\mathbf{V}^{+\prime} := (V_1^+, \ldots, V_p^+),$$

$$V_j^+(y, \mathbf{t}) := V_j(y, \mathbf{t}) - \sum_{i=1}^{n} x_{nij} + V_j(-y, \mathbf{t}), \quad 1 \le j \le p, \ y \in \mathbb{R}, \mathbf{t} \in \mathbb{R}^p.$$

In the case errors are not symmetric but i.i.d. according to a known d.f. F, so that $EV_j(y, \beta) \equiv \Sigma_i\, x_{nij}\, F(y)$, a suitable class of m.d. estimators of β is defined by the relation

$$\hat{\beta}_{\mathbf{X}} := \operatorname{argmin}\{K_{\mathbf{X}}(\mathbf{t});\ \mathbf{t} \in \mathbb{R}^p\},$$

with

(2) $$K_{\mathbf{X}}(\mathbf{t}) := \int \|\mathbf{W}(y, \mathbf{t})\|^2 \, dG(y), \qquad \mathbf{t} \in \mathbb{R}^p,$$

$$\mathbf{W}(y, \mathbf{t}) := (\mathbf{X}'\mathbf{X})^{-1/2}\{\mathbf{V}(y, \mathbf{t}) - \mathbf{X}'\mathbf{1}\, F(y)\}, \qquad y \in \mathbb{R},\ \mathbf{t} \in \mathbb{R}^p,$$

$$\mathbf{1}' := (1, \ldots, 1)_{1\mathrm{x}n}.$$

Chapter 5 discusses the existence, the asymptotic distribution, the robustness and the asymptotic optimality of $\boldsymbol{\beta}^{+}_{\mathbf{X}}$ and $\hat{\boldsymbol{\beta}}_{\mathbf{X}}$ under (1.1.1) with heteroscedastic errors. For example, if $p = 1$ in (1.1.1) and the design variable is nonnegative then the asymptotic variance of $\boldsymbol{\beta}^{+}_{\mathbf{X}}$ is smaller than that of $\boldsymbol{\beta}^{+}_{1}$ for a large class of symmetric error d.f.'s F and integrating measures G. A similar result holds about $\hat{\boldsymbol{\beta}}_{\mathbf{X}}$ and for $p \geq 1$. Chapter 5 also discusses several other m.d. estimators of $\boldsymbol{\beta}$ and their asymptotic theory under (1.1.1) with heteroscedastic errors. These include analogues of $\hat{\boldsymbol{\beta}}_{\mathbf{X}}$ when the common error d.f. is unknown and some m.d. estimators corresponding to certain supremum distances based on $\mathbf{V}$.

Closely related to the problem of minimum distance estimation is the problem of testing the goodness-of-fit hypothesis H_0: $F_{ni} \equiv F_0$, F_0 a known d.f.. One test statistic for this problem is

$$\hat{D}_1 := \sup_y |n^{1/2}\{H_n(y, \hat{\boldsymbol{\beta}}) - F_0(y)\}|,$$

where $\hat{\boldsymbol{\beta}}$ is an estimator of $\boldsymbol{\beta}$. This test statistic is suggested by looking at the estimated residuals and mimicking the one sample location model technique. In general, its large sample distribution depends on the design matrix. In addition, it does not reduce to the Kiefer (1959) tests of goodness-of-fit in the k-sample location problem when (1.1.1) is reduced to this model. Test statistics that overcome these deficiencies are

$$\hat{D}_2 := \sup_y |\mathbf{W}^0(y, \hat{\boldsymbol{\beta}})|, \qquad \hat{D}_3 := \sup_y \|\mathbf{W}^0(y, \hat{\boldsymbol{\beta}})\|,$$

where $\mathbf{W}^0$ is equal to the $\mathbf{W}$ of (2) with $F = F_0$. Another natural class of tests is based on $K^0_{\mathbf{X}}(\hat{\boldsymbol{\beta}}_{\mathbf{X}})$, where $K^0_{\mathbf{X}}$ is equals to the $K_{\mathbf{X}}$ of (2) with $\mathbf{W}$ replaced by $\mathbf{W}^0$ in there.

All of the above and several other goodness-of-fit tests are discussed at some length in Chapter 6. Section 6.2a discusses the asymptotic null distributions of the supremum distance statistics $\hat{D}_j$, $j = 1, 2, 3$. Also discussed in this section are asymptotically distribution free analogues of these tests, in a sense similar to that discussed by Durbin (1973, 1976) and Rao (1972) for the one-sample location model. Section 6.2b discusses

smooth bootstrap approximations to the null distributions of tests based on w.e.p.'s.

Tests based on L_2-distances are discussed in Section 6.3. Some modifications of goodness-of-fit tests when F_0 has a scale parameter appear in Section 6.4 while tests of the symmetry of the errors are discussed in Section 6.5.

1.4. RANDOMLY WEIGHTED EMPIRICAL PROCESSES

A ***randomly weighted empirical process*** (r.w.e.p.) corresponding to the random variables (r.v.'s) $\zeta_{n1}, \ldots, \zeta_{nn}$, the random noise $\delta_{n1}, \ldots, \delta_{nn}$ and the random real weights $h_{n1}, \ldots, h_{nn}$ is defined to be

$$(1) \qquad V_h(x) := n^{-1} \sum_{i=1}^{n} h_{ni}\, I(\zeta_{ni} \le x + \delta_{ni}), \qquad x \in \mathbb{R},\ n \ge 1.$$

Examples of r.w.e.p.'s are provided by the w.e.p.'s $\{V_j;\ 1 \le j \le p\}$ of (1.1.2) in the case the design variables are random. More importantly, r.w.e.p.'s arise naturally in autoregression models. To illustrate this, let $\mathbf{Y}_0 = (X_0, \ldots, X_{1-p})'$ be an observable random vector, $\{\epsilon_i,\ i \ge 1\}$ be i.i.d. r.v.'s, independent of $\mathbf{Y}_0$, and $\boldsymbol{\rho}' = (\rho_1, \ldots, \rho_p)$ be a p-dimensional parameter vector. In the p^{th} order autoregression (AR(p)) model one observes $\{X_i\}$ obeying the relation

$$(2) \qquad X_i = \rho_1 X_{i-1} + \ldots + \rho_p X_{i-p} + \epsilon_i, \qquad i \ge 1,\ \ \boldsymbol{\rho} \in \mathbb{R}^p.$$

Processes that play a fundamental role in the robust estimation of $\boldsymbol{\rho}$ in this model are randomly weighted residual empirical processes $\mathbf{T} = (T_1, \ldots, T_p)'$, where

$$(3) \qquad T_j(x, \mathbf{t}) := n^{-1} \sum_{i=1}^{n} g(X_{i-j})\, I(X_i \le x + \mathbf{t}'\mathbf{Y}_{i-1}), \qquad x \in \mathbb{R},\ \ \mathbf{t} \in \mathbb{R}^p,$$

$\mathbf{Y}'_{i-1} = (X_{i-1}, \ldots, X_{i-p})$, $i \ge 1$, and where g is a measurable function from $\mathbb{R}$ to $\mathbb{R}$. Clearly, for each $1 \le j \le p$, $T_j(x, \boldsymbol{\rho}+n^{-1/2}\mathbf{t})$ is an example of $V_h(x)$ with $\zeta_{ni} \equiv \epsilon_i$, $\delta_{ni} \equiv n^{-1/2}\,\mathbf{t}'\mathbf{Y}_{i-1}$ and $h_{ni} \equiv g(X_{i-j})$.

It is customary to expect that a method that works for linear regression models should have an analogue that will also work in autoregression models. Indeed the above inferential procedures based on w.e.p.'s in linear regression have perfect analogues in AR(p) models in terms of $\mathbf{T}$. The generalized M-estimators of $\boldsymbol{\rho}$ as proposed by Denby and Martin (1979) corresponding to the weight function g and the score functions ψ are given as a solution $\mathbf{t}$ of the p equations

$$\int \psi(x)\, T(dx, \mathbf{t}) = 0,$$

assuming that $E\psi(\epsilon) = 0$. Clearly, the classical least square estimator is obtained upon taking $g(x) \equiv x \equiv \psi(x)$ in these equations.

A generalized R-estimator $\hat{\rho}_R$ corresponding to a score function φ is defined by the relation

$$\text{(4)} \qquad \hat{\rho}_R := \operatorname{argmin}\, \{\|S(\mathbf{t})\|;\ \mathbf{t} \in \mathbb{R}^p\})\|,$$

where

$$S(\mathbf{t}) := \int \varphi(F_n(x, \mathbf{t}))\, T(dx, \mathbf{t}),$$

$$F_n(x, \mathbf{t}) := n^{-1} \sum_{i=1}^{n} I(X_i \le x + \mathbf{t}' Y_{i-1}), \qquad x \in \mathbb{R},\ \mathbf{t} \in \mathbb{R}^p.$$

An analogue of an R-estimator of (1.2.1) is obtained by taking $g(x) \equiv x$ in (4).

The m.d. estimators ρ_X^+ that are analogues of β_X^+ of (1.3.1) are defined as minimizers, w.r.t. $\mathbf{t} \in \mathbb{R}^p$, of

$$K(\mathbf{t}) := \sum_{j=1}^{p} \int [n^{-1/2} \sum_{i=1}^{n} X_{i-j}\{I(X_i \le x + \mathbf{t}' Y_{i-1}) - I(-X_i \le x - \mathbf{t}' Y_{i-1})\}]^2\, dG(x).$$

Observe that K involves T corresponding to $g(x) \equiv x$.

Chapter 7 discusses these and some other procedures in detail. Section 7.2 contains a result that says that the r.w.e.p.'s $\{T(x, \rho + n^{-1/2}\mathbf{t}),\ x \in \mathbb{R},\ \|\mathbf{t}\| \le B\}$ and the residual empirical processes $\{F_n(x, \rho + n^{-1/2}\mathbf{t}),\ x \in \mathbb{R},\ \|\mathbf{t}\| \le B\}$ are asymptotically uniformly linear in $\mathbf{t}$, for every $0 < B < \infty$. These results are used to investigate the asymptotic behavior of G-M and R-estimators in Sections 7.3a and 7.3b respectively. In order to carry out the rank analysis in AR(p) models, one needs a consistent estimator of $Q(f)$ where now f is the error density of $\{\epsilon_i\}$. A class of such estimators is given in Section 7.3c. A large class of m.d. estimators and their asymptotics appears in Section 7.4 whereas Section 7.5 briefly discusses some tests of goodness-of-fit hypotheses pertaining to the error d.f..

The contents of Chapter 2 are basic to those of Chapters 3, 4, and parts of Chapters 6 and 7. Sections 2.2a and 2.2b contain, respectively, proofs of the weak convergence of suitably standardized w.e.p.'s and r.w.e.p.'s to continuous Gaussian processes. Even though w.e.p.'s are a special case of r.w.e.p.'s, it is beneficial to investigate their weak convergence separately. For example, the weak convergence of U_d is obtained under a fairly general independent setup and minimal conditions on $\{d_{ni}\}$ whereas that of V_h is obtained under some hierarical dependence structure on $\{\eta_{ni}, h_{ni}, \delta_{ni}\}$ and the boundedness of the weights $\{h_{ni}\}$.

In Section 2.3, the asymptotic continuity of certain standardized w.e.p.'s is used to prove the asymptotic uniform linearity of $\mathbf{V}(., \mathbf{t})$ in $\mathbf{t}$, for $\mathbf{t}$ in certain shrinking neighborhoods of $\boldsymbol{\beta}$, under fairly general heteroscedastic errors. This result is found useful in Chapter 4 when discussing M-estimators and in Chapter 6 when discussing supremum distance test statistics for goodness-of-fit hypotheses. The asymptotic continuity is also found useful in Chapter 3 to prove various results about rank and signed rank statistics under heteroscedastic errors. The asymptotic continuity of V_h-processes is found useful in Chapter 7 when discussing the AR(p) model.

Chapter 2 concludes with results on functional and bounded laws of iterated logarithm pertaining to certain w.e.p.'s. It also includes an inequality due to Marcus and Zinn (1984) that gives an exponential bound on the tail probabilities of w.e.p.'s of independent r.v.'s. This inequality is an extension of the well celebrated Dvoretzky, Kiefer and Wolfowitz (1956) inequality for the ordinary empirical process. A result about the weak convergence of w.e.p.'s when r.v.'s are p-dimensional is also stated. These results are included for completeness, without proofs. They are not used in the subsequent sections. A martingale property of a properly centered U_d process is proved in Section 2.4.

CHAPTER 2

ASYMPTOTIC PROPERTIES OF WEIGHTED EMPIRICALS

2.1. INTRODUCTION.

Let, for each $n \geq 1$, $\eta_{n1}, \ldots, \eta_{nn}$ be independent r.v.'s taking values in [0, 1] with respective d.f.'s $G_{n1}, \ldots, G_{nn}$ and $d_{n1}, \ldots, d_{nn}$ be real numbers. Define

$$(1) \qquad W_d(t) = \sum_{i=1}^{n} d_{ni}\{I(\eta_{ni} \leq t) - G_{ni}(t)\}, \qquad 0 \leq t \leq 1.$$

Observe that both V_h of (1.4.1) and W_d belong to $\mathbb{D}[0, 1]$ for each n and for any triangular arrays $\{h_{ni}, 1 \leq i \leq n\}$ and $\{d_{ni}, 1 \leq i \leq n\}$.

In this chapter we first prove certain weak convergence results about suitably standardized W_d and V_h processes. This is done in Sections 2.2a and 2.2b, respectively. Sections 2.3.1 uses the asymptotic continuity of a certain W_d-process to obtain the asymptotic uniform linearity result about $\mathbf{V}(\cdot, \mathbf{u})$ of (1.1.2) in $\mathbf{u}$. Analogous result for $\mathbf{T}(\cdot, \mathbf{u})$ of (1.4.3) uses the asymptotic continuity of a certain V_h-process and is proved in Section 7.2.

A proof of an exponential inequality for a stopped martingale with bounded differences due to Johnson, Schechtman and Zinn (1985) and Levental (1989) is included in Section 2.2b. This inequality is of general interest and an important tool needed to carry out a chaining argument pertaining to the weak convergence of V_h.

Section 2.4 treats laws of iterated logarithm pertaining to W_d, the weak convergence of W_d when $\{\eta_i\}$ are in $[0, 1]^p$, the weak convergence of W_d w.r.t. some other metrics when $\{\eta_i\}$ are in [0, 1], an embedding result for W_d when $\{\eta_i\}$ are i.i.d. uniform [0, 1] r.v.'s, and a proof of its martingale property. It also includes an exponential inequality for the tail probabilities of w.e.p.'s of independent r.v.'s. This inequality is an extension of the well celebrated Dvoretzky, Kiefer and Wolfowitz (1956) inequality for the ordinary empirical process. These results are stated for the sake of completeness, without proofs. They are not used in the subsequent sections.

2.2. WEAK CONVERGENCE

2.2a. W_d – Processes.

In this section we give two proofs of the weak convergence of suitably standardized $\{W_d\}$ to a limit in $\mathbb{C}[0, 1]$. Accordingly, let

(1) $$G_d(t) := \sum_{i=1}^{n} d_{ni}^2\, G_{ni}(t), \qquad 0 \le t \le 1,$$

and

(2) $$\mathcal{C}_d(s, t) := \sum_{i=1}^{n} d_{ni}^2\, [G_{ni}(s \wedge t) - G_{ni}(s)\, G_{ni}(t)], \qquad 0 \le s, t \le 1.$$

Let $\mathscr{d}$ denote the *supremum metric.*

Theorem 2.2a.1. *Let* $\{\eta_{ni}\}$, $\{d_{ni}\}$ *and* $\{G_{ni}\}$ *be as in Section* 2.1. *In addition assume that the following hold:*

(N1) $$\tau_d^2 := \sum_{i=1}^{n} d_{ni}^2 = 1, \text{ for all } n \ge 1.$$

(N2) $$\max_{1 \le i \le n} d_{ni}^2 \longrightarrow 0.$$

(C) $$\lim_{\delta \to 0} \limsup_n \sup_{0 \le t \le 1-\delta} [G_d(t + \delta) - G_d(t)] = 0.$$

Then, for every $\epsilon > 0$,

(i) $$\lim_{\delta \to 0} \limsup_n P\Big(\sup_{|t-s| < \delta} |W_d(t) - W_d(s)| > \epsilon\Big) = 0.$$

(ii) *Moreover,* $W_d \Rightarrow$ *some* W *on* $(\mathbb{D}[0, 1], \mathscr{d})$ *if and only if for every* $0 \le s, t \le 1$, $\mathcal{C}_d(s, t)$ *converges to some covariance function* $\mathcal{C}(s, t)$.

In this case W *is necessarily a continuous Gaussian process with covariance function* $\mathcal{C}$ *and* $W(0) = 0 = W(1)$.

Remark 2.2a.1. Perhaps a remark about the labeling of the conditions is in order. The letter **N** in (**N1**) and (**N2**) stands for **Noether** who was the first person to use these conditions to obtain the asymptotic normality of certain weighted sums of r.v.'s. See Noether (1949).

The letter **C** in the condition (**C**) stands for the specified *continuity* of the sequence $\{G_d\}$. Observe that the d.f.'s $\{G_i\}$ need not be continuous for each i and n; only $\{G_d\}$ needs to be equicontinuous in the sense of (**C**). Of course if $\{\eta_i\}$ are i.i.d. G then, because of (**N1**), (**C**) is equivalent to the continuity of G. □

The proof of the theorem will follow from the following *two* lemmas.

Lemma 2.2a.1. *For any* $0 \le s \le t \le u \le 1$ *and each* $n \ge 1$

$$E|W_d(t) - W_d(s)|^2 |W_d(u) - W_d(t)|^2$$

(3a) $$\le 3\, [G_d(u) - G_d(t)][G_d(t) - G_d(s)].$$

$$\text{(3b)} \qquad \leq 3[G_d(u) - G_d(s)]^2.$$

Proof. Fix $0 \leq s, t, u \leq 1$ and let

$$p_i = G_i(t) - G_i(s), \qquad q_i = G_i(u) - G_i(t),$$

$$\alpha_i = I(s < \eta_i \leq t) - p_i, \qquad \beta_i = I(t < \eta_i \leq u) - q_i, \quad 1 \leq i \leq n.$$

Observe that $E\,\alpha_i = 0 = E\,\beta_j$ for all $1 \leq i,\ j \leq n$, $\{\alpha_i\}$ are independent as are $\{\beta_j\}$ and that α_i is independent of β_j for $i \neq j$. Moreover,

$$W_d(t) - W_d(s) = \Sigma_i\, d_i\alpha_i\,, \qquad W_d(u) - W_d(t) = \Sigma_i\, d_i\beta_i.$$

Now expand and multiply the quadratics and use the above facts to obtain

$$\text{(4)} \qquad E|W_d(t) - W_d(s)|^2 |W_d(u) - W_d(t)|^2$$

$$= \sum_i d_i^4\, E\alpha_i^2\beta_i^2 + \sum_{i \neq j}\sum d_i^2\, d_j^2\, E\alpha_i^2\, E\beta_j^2 + 2 \sum_{i \neq j}\sum d_i^2\, d_j^2\, E(\alpha_i\beta_i)\, E(\alpha_j\, \beta_j).$$

But

$$E\alpha_i^2 = p_i(1 - p_i), \qquad E\beta_j^2 = q_j(1 - q_j),$$

$$\begin{aligned} E\alpha_i^2\beta_i^2 &= (1 - p_i)^2\, p_i q_i^2 + (1 - q_i)^2\, q_i p_i^2 + p_i q_i (1 - q_i - p_i) \\ &\leq \{(1 - p_i) + (1 - q_i) + (1 - q_i - p_i)\}\, p_i q_i \\ &\leq 3 p_i q_i, \end{aligned}$$

$$\begin{aligned} E(\alpha_i\beta_i) &= -(1 - p_i)\, p_i q_i - (1 - q_i)\, q_i p_i + p_i q_i\,(1 - q_i - p_i) \\ &= -p_i q_i\,, \qquad 1 \leq i \leq n, \quad 1 \leq j \leq n. \end{aligned}$$

Therefore,

$$\text{(5)} \qquad \text{LHS (4)} \leq 3\{\Sigma_i\, d_i^4\, p_i q_i + \sum_{i \neq j}\sum d_i^2 d_j^2\, p_i q_j\} = 3[\Sigma_i\, d_i^2 p_i]\,[\Sigma_j\, d_j^2\, q_j].$$

This completes the proof of (3a), in view of the definition of $\{p_i, q_j\}$. That of (3b) follows from (3a), (1) and the monotonicity of the G_i, $1 \leq i \leq n$. □

Lemma 2.2a.2. *For every* $\epsilon > 0$ *and* $s \leq u$,

$$\text{(6)} \qquad P[\sup_{s \leq t \leq u} |W_d(t) - W_d(s)| \geq \epsilon]$$

$$\leq \kappa\, \epsilon^{-4}\, [G_d(u) - G_d(s)]^2 + P[|W_d(u) - W_d(s)| \geq \epsilon/2]$$

where κ *does not depend on* ϵ, n *or on any underlying quantity.*

Proof. Let $\delta = u - s$, $m \geq 1$ be an integer,

$$(7) \qquad \xi_j = W_d((j/m)\delta + s) - W_d(((j-1)/m)\delta + s), \qquad 1 \leq j \leq m,$$

$$S_k = \sum_{j=1}^{k} \xi_j\,, \qquad M_m = \max_{1 \leq k \leq m} |S_k|.$$

The right continuity of W_d implies that for each n and each sample path, $M_m \longrightarrow \sup\{|W_d(t) - W_d(s)|;\ s \leq t \leq u\}$ as $m \longrightarrow \infty$, w.p.1. In view of Lemma 2.2a.1, Lemma A.1 in the Appendix is applicable to the above r.v.'s $\{\xi_j\}$ with $\gamma = 2$, $\alpha = 1$ and

$$u_j = 3^{1/2}\{G_d((j/m)\delta + s) - G_d(((j-1)/m)\delta + s)\}, \qquad 1 \leq j \leq m.$$

Hence (6) follows from that lemma and the right continuity of W_d. □

Proof of Theorem 2.2a.1. For a $\delta > 0$, let $r = [\delta^{-1}]$, the greatest integer less than or equal to $1/\delta$. Define $t_j = j\delta$, $1 \leq j \leq r$ and $t_0 = 0$. Let $\Gamma_j = W_d(t_j) - W_d(t_{j-1})$, $1 \leq j \leq r$. Then

$$P(\sup_{|t-s|<\delta} |W_d(s) - W_d(s)| \geq \epsilon)$$

$$\leq \sum_{j=1}^{r} P[\sup_{t_{j-1} \leq s \leq t_j} |W_d(s) - W_d(t_{j-1})| \geq \epsilon/3]$$

$$\leq \kappa\epsilon^{-2} \sum_{j=1}^{r} [G_d(t_j) - G_d(t_{j-1})]^2 + \sum_{j=1}^{r} P[|\Gamma_j| \geq \epsilon/6]$$

$$\leq \kappa\epsilon^{-2} \sup_{0 \leq t \leq 1-\delta} [G_d(t+\delta) - G_d(t)] + \sum_{j=1}^{r} P[|\Gamma_j| \geq \epsilon/6]$$

$$(8) \qquad = I_n(\delta) + II_n(\delta), \qquad \text{(say)}.$$

In the above the first inequality follows from Lemma A.2 of the Appendix, the second inequality follows from Lemma 2.2a.2 above and the last inequality follows because, by (N1),

$$(9) \qquad \sum_{j=1}^{r} [G_d(t_j) - G_d(t_{j-1})] \leq G_d(1) = 1.$$

Next, observe that

$$(10) \quad \sigma_j^2 := \mathrm{Var}(\Gamma_j) = \Sigma_i\, d_i^2\{G_i(t_j) - G_i(t_{j-1})\}\, \{1 - G_i(t_j) + G_i(t_{j-1})\},$$

$$\leq G_d(t_j) - G_d(t_{j-1}), \qquad 1 \leq j \leq r,$$

and, by (9), that

$$\sum_{j=1}^{r} \sigma_j^4 \le \sup_{0 \le t \le 1-\delta} [G_d(t+\delta) - G_d(t)], \quad \text{all } r \text{ and all } n. \tag{11}$$

Furthermore, (N1) and (N2) enable one to apply the Lindeberg–Feller Central Limit Theorem (L–F CLT) to conclude that $\sigma_j^{-1}\Gamma_j \underset{d}{\longrightarrow} Z$, Z a N(0, 1) r.v. Therefore, for every $\delta > 0$ (or $r < \infty$)

$$|II_n(\delta) - \sum_{j=1}^{r} P(|Z| \ge (\epsilon/6)\sigma_j^{-1})| \longrightarrow 0 \text{ as } n \longrightarrow \infty. \tag{12}$$

By the Markov Inequality applied to the summands in the second term of (12) and by (11),

$$\begin{aligned} \limsup_n II_n(\delta) &\le 3 \limsup_n \sum_{j=1}^{r} (6\sigma_j/\epsilon)^4 \qquad (EZ^4 = 3) \\ &\le \kappa\epsilon^{-4} \limsup_n \sup_{0 \le t \le 1-\delta} [G_d(t+\delta) - G_d(t)]. \end{aligned} \tag{13}$$

The result (i) now follows from (13), (8) and the assumption (C).

Proof of (ii). Suppose $\mathcal{C}_d \longrightarrow \mathcal{C}$. Let m be a positive integer, $0 \le t_1, \ldots, t_m \le 1$ and $a_1, \ldots, a_m$ be arbitrary but fixed numbers. Consider

$$T_n := \sum_{j=1}^{m} a_j W_d(t_j) = \sum_{i=1}^{n} d_i \nu_i \tag{14}$$

where

$$\nu_i := \sum_{j=1}^{m} a_j\{I(\eta_i \le t_j) - G_i(t_j)\}, \qquad 1 \le i \le n.$$

Note that

$$|\nu_i| \le \sum_{j=1}^{m} |a_j| < \infty, \qquad 1 \le i \le n. \tag{15}$$

Also, $\mathrm{Var}(T_n) \longrightarrow \sigma^2 := \sum_{j=1}^{m}\sum_{r=1}^{m} a_j a_r \mathcal{C}(t_j, t_r)$. In view of (N1) and (N2), the L–F CLT yields that $T_n \underset{d}{\longrightarrow} N(0, 1)$. Hence all finite dimensional distributions of W_d converge weakly to those of a Gaussian process W with the covariance function $\mathcal{C}$ and $W(0) = 0 = W(1)$. In view of (i), this implies that $W_d \Rightarrow W$ in $(\mathbb{D}[0, 1], \mathcal{d})$ with W denoting a continuous Gaussian process tied down at 0 and 1.

Conversely, suppose $W_d \Rightarrow W$. By (i), W is in $\mathbb{C}[0, 1]$. In particular the T_n of (14) converges in distribution to $T := \sum_{j=1}^{m} a_j W(t_j)$. Moreover, (15) and (N1) imply that, for all $n \ge 1$,

$$ET_n^4 = E(\Sigma_i d_i \nu_i)^4 = \Sigma_i d_i^4 E\nu_i^4 + 3 \sum_{i \neq j}\sum d_i^2 d_j^2 E\nu_i^2 E\nu_j^2 \leq 3(\sum_{j=1}^{m} |a_j|)^4,$$

Therefore $\{T_n^2, n \geq 1\}$ is uniformly integrable and hence

$$ET_n^2 = \sum_{j=1}^{m} \sum_{k=1}^{m} a_j a_k \mathcal{C}_d(t_j, t_k) \longrightarrow \sum_{j=1}^{m} \sum_{k=1}^{m} a_j a_k \operatorname{Cov}[W(t_j), W(t_k)]$$

for any set of numbers $0 \leq \{t_j\} \leq 1$ and any finite real numbers $a_1, \dots, a_m$. Hence

$$\mathcal{C}_d(s, t) \longrightarrow \operatorname{Cov}[W(s), W(t)] = \mathcal{C}(s, t) \text{ for all } 0 \leq s, t \leq 1.$$

Now repeat the above argument of the "only if" part to conclude that W must be a tied down Gaussian process in $\mathbb{C}[0, 1]$. □

Another set of sufficient conditions for the weak convergence of $\{W_d\}$ is given in the following

Theorem 2.2a.2. *Under the notation of Theorem* 2.2a.1, *suppose that* (N1) *holds. In addition, assume that the following hold:*

(B) $$n \max_{1 \leq i \leq n} d_{ni}^2 = O(1).$$

and

(D) $n^{-1} \Sigma_i G_{ni}(t) - t$ *is nonincreasing in* t, $0 \leq t \leq 1$, $n \geq 1$.

Then also (i) *and* (ii) *of Theorem* 2.2a.1 *hold.*

Remark 2.2a.1. Clearly (B) implies (N2). Moreover

$$\begin{aligned}[G_d(t + \delta) - G_d(t)] &\leq n \max_i d_i^2 [n^{-1}\Sigma_i \{G_i(t + \delta) - G_i(t)\}] \\ &= n \max_i d_i^2 [n^{-1}\Sigma_i \{G_i(t + \delta) - (t + \delta)\} \\ &\qquad - n^{-1}\Sigma_i \{G_i(t) - t\} + \delta] \\ &\leq n \max_i d_i^2 \delta, \qquad 0 \leq t \leq 1 - \delta, \qquad \text{by (D)}.\end{aligned}$$

Thus (B) and (D) together imply (N2) and (C). Hence Theorem 2.2a.2 follows from Theorem 2.2a.1. However, we can also give a different proof of Theorem 2.2a.2 which is direct and quite interesting (see (19) below). This proof will be based on the following *three* lemmas.

Lemma 2.2a.3. *Under* (D), *for all* $n \geq 1$,

(16) $E|W_d(t) - W_d(s)|^4 \le k_d^2\,\{3(t-s)^2 + (t-s)n^{-1}\}, \qquad 0 \le s, t \le 1$

where $k_d^2 := n \max_{1\le i\le n} d_{ni}^2$.

Proof. Suppose $0 \le s \le t \le 1$. Let α_i and p_i be as in the proof of Lemma 2.2a.1. Using the independence of $\{\alpha_i\}$ and the fact that $E\alpha_i = 0$ for all $1 \le i \le n$, one obtains

$$\begin{aligned} E|W_d(t) - W_d(s)|^4 &= E(\Sigma d_i\,\alpha_i)^4 \\ &= \Sigma_i\, d_i^4\, E\alpha_i^4 + 3\,\Sigma\Sigma_{i\ne j}\, d_i^2\, d_j^2\, E\alpha_i^2\, E\alpha_j^2 \\ &= \Sigma_i\, d_i^4\,\{E\alpha_i^4 - 3\,E^2(\alpha_i^2)\} + 3(\Sigma_i\, d_i^2\, E\alpha_i^2)^2 \\ &= \Sigma_i\, d_i^4\, p_i(1-p_i)\,(1 - 6p_i(1-p_i)) + \\ &\qquad + 3\,[\Sigma_i\, d_i^2\, p_i\,(1-p_i)]^2 \end{aligned}$$

(17) $$\le k_d^2\,\{n^{-2}\Sigma_i\, p_i + 3(n^{-1}\Sigma_i\, p_i)^2\}.$$

But $s \le t$ and (D) imply

$$0 \le n^{-1}\Sigma_i\, p_i = n^{-1}\Sigma_i\,[G_i(t) - G_i(s)] \le (t-s).$$

Hence,

$$\text{l.h.s. } (16) \le k_d^2\,\{n^{-1}(t-s) + 3(t-s)^2\}, \qquad 0 \le s \le t \le 1.$$

The proof is completed by interchanging the role of s and t in the above argument in the case $t \le s$. □

Next, define, for $(i-1)/n \le t \le i/n, \quad 1 \le i \le n$,

(18) $$Z_d(t) = W_d((i-1)/n) + \{nt - (i-1)\}\,[W_d(i/n - W_d(i-1)/n].$$

Lemma 2.2a.4. *The assumption* (D) *implies that*

(19) $$E|Z_d(t) - Z_d(s)|^4 \le k_d^2\, 144\,|t-s|^2, \quad 0 \le s, t \le 1, \ n \ge 1.$$

If, in addition, (N1) *and* (B) *hold, then,*

(20) $$\sup_t |W_d(t) - Z_d(t)| = o_p(1).$$

Proof. Let $n \ge 1$ and $0 \le s, \ t \le 1$ be arbitrary but fixed. Choose integers $1 \le i, \ j \le n$ such that

(21) $\qquad (i-1)/n \le s \le i/n \;$ and $\; (j-1)/n \le t \le j/n.$

For the sake of convenience, let

$$\delta_{k,m} := |Z_d(m/n) - Z_d(k/n)| = |W_d(m/n) - W_d(k/n)|,$$

$$b_{k,m} := 4k_d^2\,[(m-k)/n]^2, \qquad m,\ k \text{ integers;}$$

$$\Delta_{u,v} := |Z_d(u) - Z_d(v)|, \qquad 0 \le u,\ v \le 1.$$

From (16),

(22) $\qquad E\delta_{k,m}^4 \le k_d^2\,\{3(m-k)^2/n^2 + n^{-2}\cdot|m-k|\} \le 4k_d^2\,[(m-k)/n]^2 = b_{k,m}.$

The proof of (19) will be completed by considering the following three cases.

Case 1. $i < j-1$. Then because of (18) and (21),

$$\Delta_{s,t} \le \max\{\delta_{i,j-1},\ \delta_{i,j},\ \delta_{i-1,j-1},\ \delta_{i-1,j}\}$$

which entails that

(23)
$$\begin{aligned} E\,\Delta_{s,t}^4 &\le E\{\delta_{i,j-1}^4 + \delta_{i,j}^4 + \delta_{i-1,j-1}^4 + \delta_{i-1,j}^4\} \\ &\le b_{i,j-1} + b_{i,j} + b_{i-1,j-1} + b_{i-1,j} \qquad \text{(by (21))} \\ &\le 4\,b_{i-1,j} = 16\,k_d^2\,[(j-(i-1))/n]^2 \end{aligned}$$

where the last inequality follows from $0 \le j-i-1 < j-i < j-(i-1)$.

Note that (21), $i < j-1$ and i, j integers imply that

(24) $$3(t-s) \ge [j-(i-1)]/n.$$

From (23) and (24) one obtains

(25) $$E\Delta_{s,t}^4 \le 144\,k_d^2\,(t-s)^2.$$

Case 2. $i = j$. In this case $(i-1)/n \le s,\ t \le i/n$. From (18) one has

$$\Delta_{s,t} = n|t-s|\ \delta_{i-1,i}$$

so that from (22)

(26) $$E\Delta_{s,t}^4 < n^4(t-s)^4 \cdot 4k_d^2 \cdot n^{-2} \le 4k_d^2\,(t-s)^2.$$

The last inequality follows because $n(t-s) \le 1$.

Case 3. $i = j-1$. By the triangle inequality

$$\Delta_{s,t} \le 2\max(\Delta_{s,i/n}, \Delta_{i/n,t}).$$

Thus by Case 2, applied once with s and i/n and once with i/n and t, one obtains

$$E\,\Delta_{s,t}^4 \le 2^4\{E\Delta_{s,i/n}^4 + E\Delta_{i/n,t}^4\}$$

(27) $$\le 2^6\,k_d^2\,\{(i/n-s)^2 + (t-i/n)^2\} \le 2^7\,k_d^2\,(t-s)^2.$$

In view of (27), (26) and (25), the proof of (19) is complete.

To **prove** (20), let $d_{i+} = \max(0, d_i)$, $d_{i-} = \max(0, -d_i)$. Then one has $d_i = d_{i+} - d_{i-}$. Decompose W_d and Z_d accordingly. Note that $\max(d_{i+}^2, d_{i-}^2) = d_{i+}^2 + d_{i-}^2 = d_i^2$, $1 \le i \le n$. This and (N1) imply that $\tau_{d+} \le 1$, $\tau_{d-} \le 1$. It also implies that if (N2) is satisfied by the $\{d_i\}$ then it is also satisfied by $\{d_{i+}, d_{i-}\}$. By the triangle inequality,

(28) $$\|W_d - Z_d\|_\infty \le \|W_{d+} - Z_{d+}\|_\infty + \|W_{d-} - Z_{d-}\|_\infty.$$

Moreover $d_{i+} \wedge d_{i-} \ge 0$, for all i. Therefore, it is enough to prove (20) for $d_i \ge 0$, $1 \le i \le n$. Accordingly suppose that is the case. Then

(29) $$\|W_d - Z_d\|_\infty \le \mathcal{U}_1 + \mathcal{U}_2,$$

where

(30) $$\mathcal{U}_1 = \max_{1\le i\le n}\ \sup_{(i-1)/n\le t\le i/n} |W_d(t) - W_d((i-1)/n)|,$$

$$\mathcal{U}_2 = \max_{1\le i\le n}\ \sup_{(i-1)/n\le t\le i/n} |W_d(t) - W_d(i/n)|.$$

For $(i-1)/n \le t \le i/n$, and $d_i \ge 0$, $1 \le i \le n$,

$$\begin{aligned}|W_d(t) - W_d(i/n)| &\le |\Sigma_j\, d_j\, I(t < \eta_j \le i/n)| + \Sigma_j\, d_j\,[G_j(i/n) - G_j(t)]\\ &\le |W_d(i/n) - W_d((i-1)/n)| + \\ &\qquad + 2\,\Sigma_j\, d_j\,[G_j(i/n) - G_j((i-1)/n)]\end{aligned}$$

(31) $$\le \delta_{i-1,i} + 2\max_j d_j, \qquad \text{by (D)}.$$

Therefore, by (22), (30), (31) and the Markov inequality, for every $\epsilon > 0$ and for n sufficiently large such that $2 \max_j d_j < \epsilon$, the existence of which is guarenteed by (B),

$$P(\mathcal{U}_2 \geq \epsilon) \leq P(\max_i \delta_{i-1,i} \geq \epsilon - 2\max_i d_i)$$

$$\leq (\epsilon - 2\max_i d_i)^{-4} \sum_{i=1}^{n} E\delta_{i-1,i}^4$$

$$\leq (\epsilon - 2\max_i d_i)^{-4} \cdot 4\, k_d^2\, n^{-1} \longrightarrow 0. \tag{32}$$

Exactly similar calculations show that $\mathcal{U}_1 = o_p(1)$. □

Proof of Theorem 2.2a.2. Observe that $Z_d(0) = 0 = Z_d(1)$ and that $Z_d \in \mathbb{C}[0, 1]$ for every $n \geq 1$ and each sequence $\{d_i\}$. Hence by (19) and Theorem A.2 of the Appendix, $\{Z_d\}$ is tight in $\mathbb{C}[0, 1]$. Thus claim (i) follows from (20). To prove (ii) just argue as in the proof of (ii) of Theorem 2.2a.1 above. □

The following corollary will be useful later on. To state it we need some more notation. Let $F_{n1}, \ldots, F_{nn}$ be d.f.'s on $\mathbb{R}$ and X_{ni} be a r.v. with d.f. F_{ni}, $1 \leq i \leq n$. Define

$$H(x) := n^{-1} \Sigma_i F_{ni}(x),\ x \in \mathbb{R}; \qquad H^{-1}(t) := \inf\{x; H(x) \geq t\},\ 0 \leq t \leq 1; \tag{33}$$

$$L_{ni}(t) := F_{ni}(H^{-1}(t)),\quad 1 \leq i \leq n; \qquad L_d(t) := \Sigma_i d_{ni}^2 L_{ni}(t),$$

$$W_d^*(t) := \Sigma_i d_{ni}\{I(X_{ni} \leq H^{-1}(t)) - L_{ni}(t)\}, \qquad 0 \leq t \leq 1.$$

Corollary 2.2a.1. *Assume that*

(34) $X_{n1}, \ldots, X_{nn}$ *are independent r.v.'s with respective d.f.'s* $F_{n1}, \ldots, F_{nn}$ *on* $\mathbb{R}$.

In addition, suppose that $\{d_{ni}\}$, $\{F_{ni}\}$ *satisfy* (N1), (N2) *and*

$$\lim_{\delta \to 0} \limsup_n \sup_{0 \leq t \leq 1-\delta} [L_d(t + \delta) - L_d(t)] = 0. \tag{C*}$$

Then, for every $\epsilon > 0$,

$$\lim_{\delta \to 0} \limsup_n P\big(\sup_{|t-s|<\delta} |W_d^*(t) - W_d^*(s)| \geq \epsilon\big) = 0. \tag{35}$$

Proof. Follows from Theorem 2.2a.1(i) applied to $\eta_i \equiv H(X_i)$, $G_i \equiv L_i$, $1 \leq i \leq n$. □

Remark 2.2a.3. Note that if H is continuous then $n^{-1} \Sigma_i L_{ni}(t) \equiv t$. Therefore,

$$(36) \qquad \sup_{0 \le t \le 1-\delta} [L_d(t+\delta) - L_d(t)] \le n \max_i d_{ni}^2 \, \delta.$$

Thus, if we strengthen (N2) to require (B) then (C*) is *a priori* satisfied. That is, the conditions of Theorem 2.2a.2(i) are satisfied.

If $F_{ni} \equiv F$, F a continuous d.f. then $L_{ni}(t) \equiv t$. Therefore, in view of (N1), (C*) is *a priori* satisfied. Moreover $\mathcal{C}_d^*(s, t) := \mathrm{Cov}(W_d^*(s), W_d^*(t)) = s(1-t)$, $0 \le s \le t \le 1$. Therefore we obtain

Corollary 2.2a.2. *Suppose that* $X_{n1}, \ldots, X_{nn}$ *are i.i.d.* F, F *a continuous d.f.. Suppose that* $\{d_{ni}\}$ *satisfy* (N1) *and* (N2). *Then* $W_d^* \Rightarrow B$ *in* $(\mathbb{D}[0,1], \mathcal{d})$ *with* B *a Brownian bridge in* $\mathbb{C}[0,1]$. □

Observe that $d_{ni} \equiv n^{-1/2}$ satisfy (N1) and (N2). In other words the above corollary includes the well celebrated result, v.i.z., the weak convergence of the sequence of the ordinary empirical processes.

Note. A variant of Theorem 2.2a.1 was first proved in Koul (1970). The above formulation and proof is based on this work and that of Withers (1975). Theorem 2.2a.2 is motivated by the work of Shorack (1973) which deals only with the weak convergence of the W_1-process, the process W_d with $d_{ni} \equiv n^{-1/2}$. The sufficiency of condition (D) for (16) was observed by Eyster (1977). □

2.2b. V_h-processes

In this subsection we shall investigate the weak convergence of the r.w.e.p.'s $\{V_h(x),\ x \in \mathbb{R}\}$ of (1.4.1). To state the general result we need some more structure on the underlying r.v.'s.

Accordingly, let $(\Omega, \mathcal{A}, P)$ be a probability space and G be a d.f. on $\mathbb{R}$. For each integer $n \ge 1$, let $(\zeta_{ni}, h_{ni}, \delta_{ni})$, $1 \le i \le n$, be an array of trivariate r.v.'s defined on $(\Omega, \mathcal{A})$ such that $\{\zeta_{ni}, 1 \le i \le n\}$ are i.i.d. G r.v.'s and ζ_{ni} is independent of (h_{ni}, δ_{ni}) for each $1 \le i \le n$. Furthermore, let $\{\mathcal{A}_{ni}\}$ be an array of sub σ–fields such that $\mathcal{A}_{ni} \subset \mathcal{A}_{n,i+1}$, $\mathcal{A}_{ni} \subset \mathcal{A}_{n+1,i}$, $1 \le i \le n$, $n \ge 1$; (h_{n1}, δ_{n1}) is $\mathcal{A}_{n1}$-measurable; the r.v.'s $\{\zeta_{n1}, \ldots, \zeta_{n,j-1}; (h_{ni}, \delta_{ni}), 1 \le i \le j\}$ are $\mathcal{A}_{nj}$-measurable, $2 \le j \le n$; and ζ_{nj} is independent of $\mathcal{A}_{nj}$, $1 \le j \le n$. Define

$$(1) \qquad V_h(x) := n^{-1} \sum_{i=1}^n h_{ni}\, I(\zeta_{ni} \le x + \delta_{ni}), \qquad V_h^*(x) := n^{-1} \sum_{i=1}^n h_{ni}\, I(\zeta_{ni} \le x),$$

$$J_h(x) := n^{-1} \sum_{i=1}^n E[\{h_{ni}\, I(\zeta_{ni} \le x + \delta_{ni})\} \,|\, \mathcal{A}_{ni}] = n^{-1} \sum_{i=1}^n h_{ni}\, G(x + \delta_{ni}),$$

$$J_h^*(x) := n^{-1} \sum_{i=1}^n h_{ni}\, G(x),$$

$$U_h(x) := n^{1/2}[V_h(x) - J_h(x)], \quad U_h^*(x) = n^{1/2}[V_h^*(x) - J_h^*(x)], \quad x \in \mathbb{R}.$$

We are now ready to state the following

Theorem 2.2b.1. *In addition to the above, assume that the following conditions hold:*

(A1) $\sup_{n\geq 1} \max_i |h_{ni}| \leq c$, a.s., *for some constant* $c < \infty$.

(A2) $\max_i |\delta_{ni}| = o_p(1)$.

(A3) $n^{-1/2} \Sigma_i |h_{ni}\, \delta_{ni}| = O_p(1)$.

(A4) G *has a uniformly continuous density* g *w.r.t.* λ, *and* $g > 0$, *a.e.*

Then

(2) $$\| U_h - U_h^* \|_\infty = o_p(1).$$

If, in addition,

(A5) $n^{-1} \Sigma_i |h_{ni}|^2 \longrightarrow \alpha^2$ *in probability,* α *a r.v.,*

then

(3) $$U_h \Rightarrow \alpha\cdot B(G), \qquad U_h^* \Rightarrow \alpha \cdot B(G)$$

where B *is a Brownian Bridge in* $\mathbb{C}[0, 1]$, *independent of* α.

The proof of (2) uses a restricted chaining argument and an exponential inequality for martingales with bounded differences. It will be a consequence of the following *two* lemmas.

Lemma 2.2b.1. *Under* (A1) – (A4), $\forall\, \epsilon > 0$ *and for* $r = 1, 2$,

$$\lim_n P\Big(\sup_{x,y} n^{-1/2} \sum_{i=1}^{n} |h_{ni}|^r \, |G(y + \delta_{ni}) - G(x + \delta_{ni})| \leq 2c^r \epsilon\Big) = 1,$$

where the supremum is taken over the set $\{x, y \in \mathbb{R};\ n^{1/2}|G(x) - G(y)| \leq \epsilon\}$.

Proof. Let $\epsilon > 0$, $q(u) := g(G^{-1}(u))$, $0 \leq u \leq 1$; $\quad \gamma_n := \max_i |\delta_i|$,

$$\omega_n := \sup\{|q(u) - q(v)|;\ |u - v| \leq \epsilon n^{-1/2}\}$$
$$= \sup\{|g(x) - g(y)|;\ |G(x) - G(y)| \leq \epsilon n^{-1/2}\},$$
$$\Delta_n := \sup\{|g(x) - g(y)|;\ |y - x| \leq \gamma_n\}.$$

By (A4), q is uniformly continuous on [0, 1]. Hence, by (A2),

(4) $$\Delta_n = o_p(1), \qquad \omega_n = o(1).$$

But

$$\sup_{x,y} n^{-1/2} \sum_{i=1}^{n} |h_i|^r \, |G(y + \delta_i) - G(x + \delta_i)|$$
$$\le \sup_{x,y} n^{-1/2} \sum_{i=1}^{n} |h_i|^r \, |G(y) - G(x)| +$$
$$+ n^{-1/2} \sum_{i=1}^{n} |h_i|^r |\delta_i| \, [\omega_n + 2\Delta_n]$$
$$\le c^r \epsilon + O_p(1) \cdot o_p(1), \qquad \text{by (A3) and (4).}$$

This completes the proof of the Lemma. □

Lemma 2.2b.2. *Let* $\{\mathcal{F}_i,\ i \ge 0\}$ *be an increasing sequence of* σ*-fields,* m *be a positive integer,* $\tau \le m$ *be a stopping time relative to* $\{\mathcal{F}_i\}$ *and* $\{\xi_i,\ 1 \le i \le m\}$ *be a sequence of real valued martingale differences w.r.t.* $\{\mathcal{F}_i\}$. *In addition, suppose that*

(i) $|\xi_i| \le M < \infty$, *for some constant* $M < \infty$, $1 \le i \le m$,

and

(ii) $\sum_{i=1}^{\tau} E(\xi_i^2 | \mathcal{F}_{i-1}) \le L$, *for a constant* $L < \infty$.

Then, for every $a > 0$,

(5) $$P\Big(\Big|\sum_{i=1}^{\tau} \xi_i\Big| > a\Big) \le 2 \exp\{-(a/2M) \operatorname{arcsinh}(Ma/2L)\}.$$

Proof. Write $\sigma_i^2 = E(\xi_i^2 | \mathcal{F}_{i-1})$, $i \ge 1$. First, consider the

case $\tau = m$:

Recall the following elementary facts: For all $x \in \mathbb{R}$,

(6a) $\exp(x) - x - 1 \le 2(\cosh x - 1) \le x \sinh x$,

(6b) $(\sinh x)/x$ is increasing in $|x|$,

(6c) $x \le \exp(x - 1)$.

Because $E(\xi_i | \mathcal{F}_{i-1}) \equiv 0$ and by (i), for a $\delta > 0$ and for all $1 \le i \le m$,

$$E\{[\exp(\delta\xi_i) - 1] \mid \mathcal{F}_{i-1}\} \le E\{\delta\xi_i \sinh(\delta\xi_i) \mid \mathcal{F}_{i-1}\}, \qquad \text{by (6a)}$$

$$\le \sigma_i^2\, \delta \sinh(\delta M)/M, \qquad \text{by (6b)}. \tag{7}$$

Use a conditioning argument to obtain

$$\begin{aligned}
&E\exp\{\delta \sum_{i=1}^{m} \xi_i\} \\
&\quad = E[\exp(\delta \sum_{i=1}^{m-1} \xi_i)\, E\{\exp(\delta\, \xi_m) \mid \mathcal{F}_{m-1}\}] \\
&\quad \le E[\exp(\delta \sum_{i=1}^{m-1} \xi_i) \exp(E\{\exp(\delta\, \xi_m) \mid \mathcal{F}_{m-1}\} - 1)], && \text{by (6c)} \\
&\quad \le E[\exp(\delta \sum_{i=1}^{m-1} \xi_i) \exp(\sigma_m^2 \cdot \delta/M \cdot \sinh(\delta M))], && \text{by (7)} \\
&\quad \le E[\exp(\delta \sum_{i=1}^{m-1} \xi_i) \exp\{(L - \sum_{i=1}^{m-1} \sigma_i^2) \cdot \delta/M \cdot \sinh(\delta M)\}].
\end{aligned}$$

Observe that $L - \sum_{i=1}^{j-1} \sigma_i^2$ is $\mathcal{F}_{j-2}$ measurable, for all $j \ge 2$. Hence, iterating the above argument $m - 1$ times will give

$$E\exp\{\delta \sum_{i=1}^{m} \xi_i\} \le \exp\{L \cdot \delta/M \cdot \sinh(\delta M)\}.$$

Now, by the Markov inequality, $\forall\ a > 0$,

$$P(\sum_{i=1}^{m} \xi_i \ge a) \le E\exp\{\delta(\sum_{i=1}^{m} \xi_i - a)\} \le \exp\{\delta\,[L/M \cdot \sinh(\delta M) - a]\}.$$

The choice of $\delta = (1/M)\operatorname{arcsinh}(Ma/2L)$ in this leads to the inequality

$$P(\sum_{i=1}^{m} \xi_i \ge a) \le \exp\{(-a/2M)\operatorname{arcsinh}(Ma/2L)\}.$$

An application of this inequality to $\{-\xi_i\}$ will yield the same bound for $P(\sum_{i=1}^{m} \xi_i \le -a)$, thereby completing the proof of (5) in the case $\tau = m$. Now consider the

general case $\tau \le m$:

Let $\chi_j = \xi_j I(j \le \tau)$. Because the event $[j \le \tau] \in \mathcal{F}_{j-1}$, it follows that $\{\chi_j, \mathcal{F}_j\}$ satisfy the conditions of the previous case. Hence,

$$P(|\sum_{i=1}^{\tau} \xi_i| \ge a) = P(|\sum_{i=1}^{m} \chi_i| \ge a) \le \exp\{(-a/2M) \operatorname{arcsinh}(Ma/2L)\}. \quad \square$$

Proof of Theorem 2.2b.1. For the clarity of the proof it is important to emphasize the dependence of various underlying processes on n. Accordingly, we shall write V_n, U_n etc. for V_h, U_h etc. in the proof.

On $\mathbb{R}$ define the metric $d(x,y) := |G(x) - G(y)|^{1/2}$. This metric makes $\mathbb{R}$ totally bounded. Thus, to prove the theorem, it suffices to prove

(a) $\forall\ y \in \mathbb{R}, \qquad |U_n(y) - U_n^*(y)| = o_p(1)$,

(b) $\forall\ \epsilon > 0\ \exists\ \delta > 0\ \ni$

(i) $\limsup_n P(\sup_{d(x,y) \le \delta} |U_n(y) - U_n(x)| > \epsilon) < \epsilon$,

(ii) $\limsup_n P(\sup_{d(x,y) \le \delta} |U_n^*(y) - U_n^*(x)| > \epsilon) < \epsilon$.

Proof of (a). The fact that $U_n - U_n^*$ is a sum of conditionally centered bounded r.v.'s yields that

$$\operatorname{Var}(U_n(y) - U_n^*(y)) \le E\, n^{-1} \Sigma_i h_i^2\, |G(y + \delta_i) - G(y)| = o(1),$$

by (A1), (A2), (A4) and the Dominated Convergence Theorem.

Proof of (b)(i). The following proof of (b)(i) uses a restricted chaining argument as discussed in Pollard (1984: p. 160–162), and the exponential inequality of Lemma 2.2b.2 above.

Fix an $\epsilon > 0$. Let $a_n := [n^{1/2}/\epsilon]$, the greatest integer less than or equal to $n^{1/2}/\epsilon$, and define the grid

$$\mathcal{H}_n := \{y_j;\ G(y_j) = j\epsilon n^{-1/2},\ 1 \le j \le a_n\}, \qquad n \ge 1.$$

Also let

$$Z_i(x) := I(\zeta_i \le x + \delta_i) - G(x + \delta_i), \quad x \in \mathbb{R}, \quad 1 \le i \le n.$$

Write $h_i \equiv h_{i+} - h_{i-}$, $h_{i+} \equiv \max(0, h_i)$, so that

$$U_n(x) = n^{-1/2} \sum_{i=1}^{n} h_{i+} Z_i(x) - n^{-1/2} \sum_{i=1}^{n} h_{i-} Z_i(x) = U_n^+(x) - U_n^-(x), \quad \text{say.}$$

Thus to prove (b)(i), by the triangle inequality, it suffices to prove it for $U_n^{\pm}$ processes. The details of the proof shall be given for the U_n^+ process only; those for the U_n^- being similar.

Next, we need to define the sequence of stopping times

$$\tau_n^+ := n \wedge \max \{k \geq 1;\ \max_{x,\, y \in \mathscr{H}_n} \frac{\sum_{i=1}^{k} (h_{i+})^2 E\{[Z_i(x) - Z_i(y)]^2 \mid \mathcal{A}_i\}}{d^2(x,y)} < 3\epsilon c^2 n\}.$$

Observe that $\tau_n^+ \leq n$. To adapt the present situation to that of the Pollard, we first prove that $P(\tau_n^+ < n) \longrightarrow 0$ (see (8) below). This allows one to work with $n^{-1/2}(\tau_n^+)^{1/2}\, U^+_{\tau_n^+}$ instead of U_n^+. By Lemma 2.2b.2 and the fact that $\operatorname{arcsinh}(x)$ is increasing and concave in x, one obtains that if x, y in $\mathscr{H}_n$ are such that $d^2(x, y) \geq t\epsilon n^{-1/2}$ then

$$P(n^{-1/2}(\tau_n^+)^{1/2} \mid U^+_{\tau_n^+}(x) - U^+_{\tau_n^+}(y) \mid \geq t)$$

$$\leq 2 \exp \{- \frac{t^2}{2c d^2(x,y)} \cdot \epsilon \operatorname{arcsinh}(1/(6\epsilon^2 c))\}, \qquad \text{for all } t > 0.$$

This enables one to carry out the chaining argument as in Pollard. What remains to be done is to connect between the points in $\mathbb{R}$ and a point in $\mathscr{H}_n$ which will be done in (9) below. We shall now prove

$$\text{(8)} \qquad P(\tau_n^+ < n) \longrightarrow 0.$$

Proof of (8). For y_j, y_k in $\mathscr{H}_n$ with $y_j < y_k$, $d^2(y_j, y_k) \geq (k-j)\epsilon n^{-1/2}$. Hence, using the fact that $(h_{i+})^2 \leq h_i^2$,

$$\sum_{i=1}^{n} (h_{i+})^2\, E\{[Z_i(y_k) - Z_i(y_j)]^2 \mid \mathcal{A}_i\} / d^2(y_i, y_k)$$

$$\leq \sum_{i=1}^{n} h_i^2\, [G(y_k + \delta_i) - G(y_j + \delta_i)]\, \{(k-j)\epsilon\}^{-1}\, n^{1/2}$$

$$\leq \{(k-j)\epsilon\}^{-1}\, n^{1/2} \sum_{i=1}^{n} h_i^2 \sum_{r=j}^{k-1} [G(y_{r+1} + \delta_i) - G(y_r + \delta_i)]$$

$$\leq \epsilon^{-1} n^{1/2} \max_{1 \leq r \leq a_n} \sum_{i=1}^{n} h_i^2 \,[G(y_{r+1} + \delta_i) - G(y_r + \delta_i)].$$

Now apply Lemma 2.2b.1 with $r = 2$ to obtain

$$P\Big(\max_{1 \leq k,\, j \leq a_n} \Big[\sum_{i=1}^{n} (h_{i+})^2\, E\{[Z_i(y_k) - Z_i(y_j)]^2 \,|\, \mathcal{A}_i\} / d^2(y_j, y_k)\Big] < 3\epsilon c^2 n\Big) \longrightarrow 1.$$

This completes the proof of (8).

Next, for each $x \in \mathbb{R}$, let y_{j_x} denote the point in $\mathcal{H}_n$ that is the closest to x in d-metric from the points in $\mathcal{H}_n$ that satisfy $y_{j_x} \leq x$. We shall now prove: $\forall\ \epsilon > 0$,

$$\text{(9)} \qquad P(\sup_x |\,\mathcal{U}_n^+(x) - \mathcal{U}_n^+(y_{j_x})| > 8c\epsilon) \longrightarrow 0.$$

Proof of (9). Now write $\mathcal{V}_n^+$, $\mathcal{J}_n^+$ for $\mathcal{V}_n$, $\mathcal{J}_n$ when $\{h_i\}$ in these quantities is replaced by $\{h_{i+}\}$.

The definition of y_{j_x}, G increasing, and the fact that $h_{i+} \leq |h_i|$ for all i, imply that

$$\sup_x |\,n^{1/2}[\mathcal{J}_n^+(x) - \mathcal{J}_n^+(y_{j_x})]\,|$$
$$\leq \max_{1 \leq j \leq a_n} n^{-1/2} \sum_{i=1}^{n} |h_i|\, [G(y_{j+1} + \delta_i) - G(y_j + \delta_i)].$$

An application of Lemma 2.2b.1 with $r = 1$ now yields that

$$\text{(10)} \qquad P(\sup_x |\,n^{1/2}[\mathcal{J}_n^+(x) - \mathcal{J}_n^+(y_{j_x})]\,| > 4c\epsilon) \longrightarrow 0.$$

But $h_{i+} \geq 0$, $1 \leq i \leq n$, implies that $\mathcal{V}_n^+$ is nondecreasing in x. Therefore, using the definition of y_{j_x},

$$n^{-1/2}[\mathcal{U}_n^+(y_{j_x-1}) - \mathcal{U}_n^+(y_{j_x})] + \mathcal{J}_n^+(y_{j_x-1}) - \mathcal{J}_n^+(y_{j_x}) =$$
$$= \mathcal{V}_n^+(y_{j_x-1}) - \mathcal{V}_n^+(y_{j_x}) \leq \mathcal{V}_n^+(x) - \mathcal{V}_n^+(y_{j_x}) \leq \mathcal{V}_n^+(y_{j_x+1}) - \mathcal{V}_n^+(y_{j_x})$$
$$= n^{-1/2}[\mathcal{U}_n^+(y_{j_x+1}) - \mathcal{U}_n^+(y_{j_x})] + \mathcal{J}_n^+(y_{j_x+1}) - \mathcal{J}_n^+(y_{j_x}).$$

Hence,

$$(11) \qquad \sup_x |n^{1/2}[V_n^+(x) - V_n^+(y_{j_x})]| \le 2 \max_{1\le j\le a_n} |U_n^+(y_{j+1}) - U_n^+(y_j)| + $$
$$+ 2 \max_{1\le j\le a_n} |n^{1/2}[J_n^+(y_{j+1}) - J_n^+(y_j)]|.$$

Thus, (9) will follow from (10), (11) and

$$(12) \qquad P(\max_{1\le j\le a_n} |U_n^+(y_{j+1}) - U_n^+(y_j)| > c\,\epsilon) \longrightarrow 0.$$

In view of (8), to prove (12), it suffices to show that

$$(13) \qquad P(\max_{1\le j\le a_n} n^{-1/2}(\tau_n^+)^{1/2} |U_{\tau_n^+}^+(y_{j+1}) - U_{\tau_n^+}^+(y_j)| > c\,\epsilon) \longrightarrow 0.$$

But,

$$\text{l.h.s.}(13) \le \sum_{j=1}^{a_n} P(|\sum_{i=1}^{\tau_n^+} h_{i+}[Z_i(y_{j+1}) - Z_i(y_j)]| > c\epsilon n^{1/2}).$$

Now apply Lemma 2.2b.2 with $\xi_i \equiv h_{i+}[Z_i(y_{j+1}) - Z_i(y_j)]$, $\mathcal{F}_{i-1} \equiv \mathcal{A}_i$, $\tau \equiv \tau_n^+$, $M = c$, $a = c\epsilon n^{1/2}$, $m = n$. By the definition of τ_n^+, $L = 3c^2\epsilon^2 n^{1/2}$. Hence by Lemma 2.2b.2,

$$P(|\sum_{i=1}^{\tau_n^+} h_{i+}[Z_i(y_{j+1}) - Z_i(y_j)]| > c\epsilon n^{1/2}) \le 2\exp[-\frac{n^{1/2}\epsilon}{2}\operatorname{arcsinh}(1/6\epsilon)].$$

Since this bound does not depend on j, it follows that

$$\text{l.h.s.}(13) \le 2\epsilon^{-1} n^{1/2} \exp[-\frac{n^{1/2}\epsilon}{2}\operatorname{arcsinh}(1/6\epsilon)] \longrightarrow 0.$$

This completes the proof of (9) for U_n^+. As mentioned earlier the proof of (9) for U_n^- is exactly similar, thereby completing the proof of (b)(i).

Adapt the above proof of (b)(i) with $\delta_i \equiv 0$ to conclude (b)(ii). *Note that* (b)(ii) *holds solely under* (A1) *and the assumption that* G *is continuous and strictly increasing; the other assumptions are not required here.* The proof of (2) is now complete.

The claim (3) follows from (1), (b)(ii) above, Lemma A.3 of the Appendix and the Cramer–Wold device. □

As noted in the proof of the above theorem, the weak convergence of U_h^* holds only under (A1), (A5) and the assumption that G is continuous and strictly increasing. For an easy reference later on we state this result as

Corollary 2.2b.1. *Let the setup of Theorem* 2.2b.1 *hold. Assume that* G *is continuous and strictly increasing and that* (A1), (A5) *hold. Then,* $U_h^* \Rightarrow \alpha \cdot B(G)$, *where* B *is a Brownian bridge in* $\mathbb{C}[0, 1]$, *independent of* α. □

Remark 2.2b.1. Consider the process $\mathcal{U}_h(t) := U_h^*(G^{-1}(t)),\ 0 \le t \le 1$. Now work with the metric $|t-s|^{1/2}$ on $[0, 1]$. Upon repeating the arguments in the proof of the above theorem, modified appropriately, one can readily conclude the following

Corollary 2.2b.2. *Let the setup of Theorem* 2.2b.1 *hold. Assume that* G *is continuous and that* (A1), (A5) *hold. Then* $\{\mathcal{U}_h\} \Rightarrow \alpha \cdot B$, *where* B *is a Brownian Bridge in* $\mathbb{C}[0, 1]$, *independent of* α. □

Remark 2.2b.2. Suppose that in Theorem 2.2a.2 the r.v.'s $\eta_{n1}, \ldots, \eta_{nn}$ are i.i.d Uniform [0, 1]. Then, upon choosing $h_{ni} \equiv n^{1/2}d_{ni}$, $\zeta_{ni} \equiv G^{-1}(\eta_{ni})$, one sees that $U_h \equiv W_d(G)$, provided G is continuous. Moreover, the condition (D) is *a priori* satisfied, (B) is equivalent to (A1) and (N1) implies (A5) trivially. Consequently, for this special setup, Theorem 2.2a.2 is a special case of Corollary 2.2b.2. But in general these two results serve different purposes. Theorem 2.2a.1 is the most general for the independent setup given there and cannot be deduced from Theorem 2.2b.1. □

Note: The inequality (5) and its proof appears in Levental (1989). See also Proposition 3.1 in Johnson, Schechtman and Zinn (1985). The proof of Theorem 2.2b.1 has its roots in Levental and Koul (1989) and Koul (1991). It was recently generalized by Koul and Ossiander (1992) to include unbounded weights. □

2.3. ASYMPTOTIC UNIFORM LINEARITY (A.U.L.) OF RESIDUAL W.E.P.'s.

In this section we shall obtain the asymptotic uniform linearity (a.u.l.) of residual w.e.p.'s. It will be observed that the asymptotic continuity property of the type specified in Theorem 2.2a.1(i) is the basic tool to obtain this result. Accordingly let $\{X_{ni}\}$, $\{F_{ni}\}$, $\{H\}$ and $\{L_{ni}\}$ be as in (2.2a.33) and define

$$
\begin{aligned}
(1)\qquad & S_d(t, \mathbf{u}) := \Sigma_i\, d_{ni}\, I(X_{ni} \le H^{-1}(t) + \mathbf{c}_{ni}'\mathbf{u}), \\
& \mu_d(t, \mathbf{u}) := \Sigma_i\, d_{ni}\, F_{ni}(H^{-1}(t) + \mathbf{c}_{ni}'\mathbf{u}), \\
& Y_d(t, \mathbf{u}) := S_d(t, \mathbf{u}) - \mu_d(t, \mathbf{u}), \qquad\qquad 0 \le t \le 1,\ \mathbf{u} \in \mathbb{R}^p,
\end{aligned}
$$

where $\{\mathbf{c}_{ni},\ 1 \le i \le n\}$ are $p \times 1$ vectors of real numbers. We also need

$$\begin{aligned}
&\text{(2a)} \qquad S_d^o(x, \mathbf{u}) := \Sigma_i\, d_{ni}\, I(X_{ni} \le x + c_{ni}'\mathbf{u}), \\
&\qquad\qquad \mu_d^o(x, \mathbf{u}) := \Sigma_i\, d_{ni}\, F_{ni}(x + c_{ni}'\mathbf{u}), \\
&\qquad\qquad Y_d^o(x, \mathbf{u}) := S_d^o(x, \mathbf{u}) - \mu_d^o(x, \mathbf{u}), \qquad -\infty \le x \le \infty, \quad \mathbf{u} \in \mathbb{R}^p.
\end{aligned}$$

Clearly, if H is strictly increasing then $S_d^o(x, \mathbf{u}) \equiv S_d(H(x), \mathbf{u})$. Similar remark applies to other functions.

Throughout the text, any w.e.p. with weights $d_{ni} \equiv n^{-1/2}$ *will be indicated by the subscript* 1. Thus, e.g., $\forall\ -\infty \le x \le \infty,\ \mathbf{u} \in \mathbb{R}^p$,

$$\begin{aligned}
&\text{(2b)} \qquad S_1^o(x, \mathbf{u}) = n^{-1/2}\, \Sigma_i\, I(X_{ni} \le x + c_{ni}'\mathbf{u}), \\
&\qquad\qquad Y_1^o(x, \mathbf{u}) = n^{-1/2}\, \Sigma_i\, \{I(X_{ni} \le x + c_{ni}'\mathbf{u}) - F_{ni}(x + c_{ni}'\mathbf{u})\}.
\end{aligned}$$

Theorem 2.3.1. *In addition to* (2.2a.34), (N1), (N2), *and* (C*) *assume that d.f.'s* $\{F_{ni},\ 1 \le i \le n\}$ *have densities* $\{f_{ni},\ 1 \le i \le n\}$ *w.r.t.* λ *such that the following hold:*

(3a) $$\lim_{\delta \to 0} \limsup_n \max_{1 \le i \le n} \sup_{|x-y| \le \delta} |f_{ni}(x) - f_{ni}(y)| = 0,$$

(3b) $$\max_{i,n} \|f_{ni}\|_\infty \le k < \infty.$$

In addition, assume that

(4) $$\max_{1 \le i \le n} \|c_{ni}\| = o(1)$$

and

(5) $$\Sigma_i\, \|d_{ni}\, c_{ni}\| = O(1).$$

Then, for every $0 < B < \infty$,

(6) $$\sup\, |S_d(t, \mathbf{u}) - S_d(t, 0) - \mathbf{u}'\Sigma_i\, d_{ni}\, c_{ni}\, q_{ni}(t)| = o_p(1),$$

where $q_{ni} := f_{ni}H^{-1}$, $1 \le i \le n$, *and the supremum is taken over* $0 \le t \le 1$, $\|\mathbf{u}\| \le B$.

Consequently, if H *is strictly increasing on* $\mathbb{R}$, *then*

(7) $$\sup\, |S_d^o(x, \mathbf{u}) - S_d^o(x, 0) - \mathbf{u}'\Sigma_i\, d_{ni}\, c_{ni}\, f_{ni}(x)| = o_p(1).$$

where the supremum is taken over $-\infty \le x \le \infty$, $\|\mathbf{u}\| \le B$.

Theorem 2.3.1 is a consequence of the following *four* lemmas. In these lemmas the setup is as in the theorem.

In what follows, $\mathcal{N}(B) = \{u \in \mathbb{R}^p;\ \|u\| \le B\}$; $\sup_{t,u}$ stands for the supremum over $0 \le t \le 1$ and $u \in \mathcal{N}(B)$, unless mentioned otherwise. Let

(8) $$\nu_d(t) := \Sigma_i\, d_{ni}\, c_{ni}\, q_{ni}(t),$$

$$R_d(t, u) := S_d(t, u) - S_d(t, 0) - u'\nu_d(t), \qquad 0 \le t \le 1,\ u \in \mathbb{R}^p.$$

Lemma 2.3.1. *Under* (3), (4) *and* (5),

(9) $$\sup_{t,u} |\mu_d(t, u) - \mu_d(t, 0) - u'\nu_d(t)| = o(1).$$

Proof. Let $\delta_n = B \max_i \|c_i\|$. By (3), $\{F_i\}$ are uniformly differentiable for sufficiently large n, uniformly in $1 \le i \le n$. Hence,

$$\text{l.h.s. (9)} \le (\Sigma_i \|d_i c_i\|) \max_i \sup_{|x-y| \le \delta_n} |f_i(x) - f_i(y)| = o(1),$$

by (3), (4) and (5). □

Lemma 2.3.2. *Under* (N1), (N2), (C*), (3), (4) *and* (5), $\forall\ u \in \mathcal{N}(B)$,

(10) $$\sup_{0 \le t \le 1} |Y_d(t, u) - Y_d(t, 0| = o_p(1).$$

Proof. Fix a $u \in \mathcal{N}(B)$. The lemma will follow if we show

(i) $$Y_d(t, u) - Y_d(t, 0) = o_p(1) \text{ for each } 0 \le t \le 1,$$

and

(ii) $\forall\ \epsilon > 0$, and for $b = u$ or $b = 0$,

$$\lim_{\delta \to 0} \limsup_n P\Big(\sup_{|t-s| < \delta} |Y_d(t, b) - Y_d(s, b)| \ge \epsilon\Big) = 0.$$

Since $Y_d(., 0) = W_d^*(.)$ of (2.2a.33), for $b = 0$, (ii) follows from (2.2a.35) of Corollary 2.2a.1.

To verify (ii) for $b = u$, take $\eta_i = H(X_i - c_i'u)$, $1 \le i \le n$, in (2.2a.1). Then $Y_d(\cdot, u) \equiv W_d(\cdot)$ of (2.2a.1) and $G_i(\cdot) = F_i(H^{-1}(\cdot) + c_i'u)$, $1 \le i \le n$. Moreover,

$$(11)\quad \sup_{0\le t\le 1-\delta} \Sigma_i d_i^2 [F_i(H^{-1}(t+\delta) + c_i'u) - F_i(H^{-1}(t) + c_i'u)]$$
$$\le 2Bk \max_i \|c_i\| + \sup_{0\le t\le 1-\delta} [L_d(t+\delta) - L_d(t)], \quad \text{by (3)}$$
$$= o(1) \text{ as } n \longrightarrow \infty, \text{ then } \delta \longrightarrow 0, \text{ by (4) and } (C^*).$$

Hence (C) is satisfied by the above $\{G_i\}$. The other conditions being (N1) and (N2) which are also assumed here, it follows that the above $\{\eta_i\}$ and $\{W_d\}$ satisfy the conditions of Theorem 2.2a.1.(i). Thus (ii) for $b = u$ follows from Theorem 2.2a.1(i). Hence (ii) is proved.

To obtain (i), note that the

$$\mathrm{Var}[Y_d(t, u) - Y_d(t, 0)] \le \Sigma_i d_i^2 \,|F_i(H^{-1}(t) + c_i'u) - F_i(H^{-1}(t))|$$
$$\le Bk \max_i \|c_i\|, \qquad \text{by (3)},$$
$$= o(1), \qquad \text{by (4)}.$$

This together with the Chebychev inequality yields (i) and hence (10). □

To state and prove the next lemma we need some more notation. Let $\kappa_{ni} = \|c_{ni}\|$, $1 \le i \le n$, and define

$$(12)\quad S_d^*(t, u, b) = \Sigma\, d_{ni}\, I(X_{ni} \le H^{-1}(t)+c_{ni}'u+b\,\kappa_{ni}),$$
$$\mu_d^*(t, u, b) = E\, S_d^*(t, u, b),$$
$$Y_d^*(t, u, b) = S_d^*(t, u, b) - \mu_d^*(t, u, b), \qquad 0 \le t \le 1,\ u \in \mathbb{R}^p,\ b \in \mathbb{R}.$$

Lemma 2.3.3. *Under* (N1), (N2), (C*), (3), (4) *and* (5), $\forall\, \epsilon > 0$, $|b| < \infty$ *and* $u \in \mathcal{N}(B)$,

$$(13)\quad \lim_{\delta\to 0} \limsup_n P\Big(\sup_{|t-s|<\delta} |Y_d^*(t, u, b) - Y_d^*(s, u, b)| \ge \epsilon\Big) = 0.$$

Proof. In Theorem 2.2a.1(i), take $\eta_i = H(X_i - c_i'u - b\kappa_i)$, $1\le i\le n$. Then $W_d(\cdot) = Y_d^*(\cdot, u, b)$ and $G_i(\cdot) = F_i(H^{-1}(\cdot)+c_i'u+b\kappa_i)$, $1\le i\le n$. Again, similar to (11),

$$\sup_{0\le t\le 1-\delta} [G_d(t+\delta) - G_d(t)] \le 2k(B+b) \max_i \|c_i\| + \sup_{0\le t\le 1-\delta} [L_d(t+\delta) - L_d(t)]$$
$$= o(1), \qquad \text{by (4) and } (C^*).$$

Hence (13) follows from Theorem 2.2a.1(i). □

Lemma 2.3.4. *Under* (N1), (N2), (C*), (3), (4) *and* (5), $\forall\ \epsilon > 0$ *there is a* $\delta > 0$ *such that for every* $\mathbf{v} \in \mathcal{N}(\mathrm{B})$,

(14) $$\limsup_n P\Big(\sup_{t,\|\mathbf{u}-\mathbf{v}\|\le\delta} |R_d(t,\mathbf{u}) - R_d(t,\mathbf{v})| \ge \epsilon\Big) = 0,$$

where R_d *is defined at* (8).

Proof. Assume, without loss of generality, that $d_i \ge 0$, $1 \le i \le n$. For, otherwise write $d_i = d_{i+} - d_{i-}$, $1 \le i \le n$, where $\{d_{i+}, d_{i-}\}$ are as in the proof of Lemma 2.2a.4. Then $S_d = \tau_{d+}S_{d+} - \tau_{d-}S_{d-}$, $R_d = \tau_{d+}R_{d+} - \tau_{d-}R_{d-}$, where $\tau^2_{d+} = \Sigma_i\,(d_{i+})^2$, $\tau^2_{d-} = \Sigma_i\,(d_{i-})^2$. In view of (N1), $\tau_{d+} \le 1$, $\tau_{d-} \le 1$. Moreover, if $\{d_i\}$ satisfy (N2) and (5) above, so do $\{d_{i+}, d_{i-}\}$ because $d^2_{i+} \vee d^2_{i-} = d^2_{i+} + d^2_{i-} = d^2_i$, $1 \le i \le n$. Hence the triangle inequality will yield (14), if proved for R_{d+} and R_{d-}. But note that $d_{i+} \wedge d_{i-} \ge 0$ for all i.

Now, $\|\mathbf{u} - \mathbf{v}\| \le \delta$ implies

(15) $$-\delta\kappa_i + \mathbf{c}_i'\mathbf{v} \le \mathbf{c}_i'\mathbf{u} \le \delta\kappa_i + \mathbf{c}_i'\mathbf{v}, \qquad \kappa_i = \|\mathbf{c}_i\|, \quad 1 \le i \le n.$$

Therefore, because $d_i \ge 0$ for all i,

(16) $$\overset{*}{S}_d(t, \mathbf{v}, -\delta) \le S_d(t, \mathbf{u}) \le \overset{*}{S}_d(t, \mathbf{v}, \delta) \text{ for all } t,$$

yielding

(17) $$\begin{aligned} L_1(t,\mathbf{u},\mathbf{v}) := \overset{*}{S}_d(t, \mathbf{v}, -\delta) - S_d(t, \mathbf{v}) - (\mathbf{u} - \mathbf{v})'\boldsymbol{\nu}_d(t) &\\ \le R_d(t, \mathbf{u}) - R_d(t, \mathbf{v}) &\\ \le \overset{*}{S}_d(t, \mathbf{v}, \delta) - S_d(t, \mathbf{v}) - (\mathbf{u} - \mathbf{v})'\boldsymbol{\nu}_d(t) &=: L_2(t,\mathbf{u},\mathbf{v}). \end{aligned}$$

We shall show that there is a $\delta > 0$ such that for every $\mathbf{v} \in \mathcal{N}(\mathrm{B})$,

(18) $$P\Big(\sup_{t,\|\mathbf{u}-\mathbf{v}\|\le\delta} |L_j(t,\mathbf{u},\mathbf{v})| \ge \epsilon\Big) = o(1), \qquad j = 1, 2.$$

We shall first prove (18) for L_2. Observe that

(19) $$\begin{aligned} |L_2(t,\mathbf{u},\mathbf{v})| \le |\overset{*}{Y}_d(t, \mathbf{v}, \delta) - \overset{*}{Y}_d(t, \mathbf{v}, 0)| &\\ + |\overset{*}{\mu}_d(t, \mathbf{v}, \delta) - \overset{*}{\mu}_d(t, \mathbf{v}, 0)| + |(\mathbf{u} - \mathbf{v})'\boldsymbol{\nu}_d(t)| & \end{aligned}$$

The Mean Value Theorem, (3), and $\|\mathbf{u} - \mathbf{v}\| \le \delta$ imply

$$(20) \qquad \sup |\overset{*}{\mu}_d(t, \mathbf{v}, \delta) - \overset{*}{\mu}_d(t, \mathbf{v}, 0)| \le \delta\, k\, \Sigma_i \|d_i c_i\|,$$

$$\sup_t |(\mathbf{u}-\mathbf{v})' \nu_d(t)| \le k\, \delta\, \Sigma_i \|d_i c_i\|.$$

Let $M(t)$ denote the first term on the r.h.s. of (19). I.e.,

$$M(t) = \overset{*}{Y}_d(t, \mathbf{v}, \delta) - \overset{*}{Y}_d(t, \mathbf{v}, 0), \qquad 0 \le t \le 1.$$

(21) **Claim:** $\sup_t |M(t)| = o_p(1).$

To begin with,

$$\begin{aligned} \text{Var}\,(M(t)) &\le \Sigma_i\, d_i^2\, [F_i(H^{-1}(t) + c_i'\mathbf{v} + \delta\kappa_i) - F_i(H^{-1}(t) + c_i'\mathbf{v})] \\ &\le \delta k \max_i \kappa_i, && \text{by (3a), (3b),} \\ &= o(1), && \text{by (5).} \end{aligned}$$

Hence

$$(22) \qquad M(t) = o_p(1) \quad \text{for every } 0 \le t \le 1.$$

Next, note that, for a $\gamma > 0$,

$$\begin{aligned} \sup_{|t-s|<\gamma} |M(t) - M(s)| \le \sup_{|t-s|<\gamma} &|\overset{*}{Y}_d(t, \mathbf{v}, \delta) - \overset{*}{Y}_d(s, \mathbf{v}, \delta)| + \\ &+ \sup_{|t-s|<\gamma} |\overset{*}{Y}_d(t, \mathbf{v}, 0) - \overset{*}{Y}_d(s, \mathbf{v}, 0)|. \end{aligned}$$

Apply Lemma 2.3.3 twice, once with $b = \delta$ and once with $b = 0$, to obtain that $\forall\ \epsilon > 0$,

$$(23) \qquad \lim_{\gamma \to 0} \limsup_n P\Big(\sup_{|t-s|<\gamma} |M(t) - M(s)| \ge \epsilon \Big) = 0.$$

But (23) and (22) imply the Claim (21).

Now choose $\delta > 0$ so that

$$(24) \qquad \limsup_n \delta\, k\, \Sigma_i \|d_i c_i\| \le \epsilon/3. \qquad \text{(use (5) here).}$$

From (19), (20) and (21) one readily obtains

$$\limsup_n P\Big(\sup_{t,\|u-v\|\le\delta} |L_2(t,u,v)| \ge \epsilon\Big) \le \limsup_n P\big(\sup_t |M(t)| \ge \epsilon/3\big) = 0.$$

This prove (18) for L_2. A similar argument proves (18) for L_1 with the same δ as in (24), thereby completing the proof of the Lemma. □

Proof of Theorem 2.3.1. Fix an $\epsilon > 0$ and choose a $\delta > 0$ satisfying (24). By the compactness of $\mathcal{N}(B)$ there exist points $v_1, \dots, v_r$ in $\mathcal{N}(B)$ such that for any $u \in \mathcal{N}(B)$, $\|u - v_j\| \le \delta$ for some $j = 1, 2, \dots, r$. Thus

$$\begin{aligned}
&\limsup_n P\big(\sup_{t,u} |R_d(t, u)| \ge \epsilon\big) \\
&\le \sum_{j=1}^{r} \limsup_n P\Big(\sup_{t,\|u-v_j\|\le\delta} |R_d(t, u) - R_d(t, v_j)| \ge \epsilon/2\Big) \\
&\qquad + \sum_{j=1}^{r} \limsup_n P\big(\sup_t |R_d(t, v_j)| \ge \epsilon/2\big) = 0
\end{aligned}$$

by Lemmas 2.3.2 and 2.3.4. □

Remark 2.3.1. Upon a reexamination of the above proof one finds that Theorem 2.3.1 is a sole consequence of the continuity of certain w.e.p.'s and the smoothness of $\{F_{ni}\}$. Note that the above proof does not use the full force of the weak convergence of these w.e.p.'s. □

Remark 2.3.2. By the relationship

$$R_d(t, u) = \mathcal{Y}_d(t, u) - \mathcal{Y}_d(t, 0) + \mu_d(t, u) - \mu_d(t, 0) - u'\nu_d(t)$$

and by Lemma 2.3.1, (6) of Theorem 2.3.1 is equivalent to

$$\sup_{0\le t\le 1,\, u\in\mathcal{N}(B)} |\mathcal{Y}_d(t, u) - \mathcal{Y}_d(t, 0)| = o_p(1). \tag{25}$$

This will be useful when dealing with w.e.p.'s based on ranks in Chapter 3. □

The above theorem needs to be extended and reformulated when dealing with a linear regression model with an unknown scale parameter or with M-estimators in the presence of a preliminary scale estimator. To that end, define, for $x, s \in \mathbb{R}$, $0 \le t \le 1$, $u \in \mathbb{R}^p$,

$$\begin{aligned}
(26)\qquad S_d(s, t, u) &:= \Sigma_i\, d_{ni} I(X_{ni} \le (1+sn^{-1/2})H^{-1}(t) + c_{ni}'u), \\
S_d^o(s, x, u) &:= \Sigma_i\, d_{ni} I(X_{ni} \le (1+sn^{-1/2})x + c_{ni}'u),
\end{aligned}$$

and define $\mathcal{Y}_d(s, t, u)$, $\mu_d(s, t, u)$ similarly. We are now ready to prove

Theorem 2.3.2. *In addition to the assumptions of Theorem* 2.3.1, *assume that*

$$(27)\qquad \max_{i,n} \sup_x |x\, f_{ni}(x)| \le k < \infty.$$

Then

$$(28)\qquad \sup |S_d(s, t, \mathbf{u}) - S_d(0, t, 0) - \Sigma_i\, d_{ni}\{sn^{-1/2}H^{-1}(t) + \mathbf{c}_{ni}'\mathbf{u}\}q_{ni}(t)| = o_p(1).$$

where the supremum is taken over $|s| \le b$, $\mathbf{u}\in\mathcal{N}(B)$, $0 \le t \le 1$.

Consequently, if H *is strictly increasing for all* $n \ge 1$, *then*

$$(29)\qquad \sup|S_d^o(s, x, \mathbf{u}) - S_d^o(0, x, 0) - \Sigma d_{ni}\, \{sn^{-1/2}x + \mathbf{c}_{ni}'\mathbf{u}\}f_{ni}(x)| = o_p(1).$$

where the supremum is taken over $|s| \le b$, $\mathbf{u}\in\mathcal{N}(B)$ *and* $x\in\mathbb{R}$.

Sketch of proof. The argument is quite similar to that of Theorem 2.3.1. We briefly indicate the modifications of the previous proof.

An analogue of Lemma 2.3.1 will now assert

$$\sup|\mu_d(s, t, \mathbf{u}) - \mu_d(1, t, 0) - \{(n^{-1/2}\Sigma d_i q_i(t)H^{-1}(t))s + \mathbf{u}'\nu_d(t)\}| = o(1).$$

This uses (3), (4), (5), (27) and (N1).

An analogue of Lemma 2.3.2 is obtained by applying Theorem 2.2a.1(i) to $\eta_i := H((X_i - \mathbf{c}_i'\mathbf{u})\sigma_n^{-1})$, $1\le i\le n$, $\sigma_n := (1+sn^{-1/2})$. This states that for every $|s| \le b$ and every $\mathbf{u}\in\mathcal{N}(B)$,

$$(30)\qquad \sup_{0\le t\le 1} |Y_d(s, t, \mathbf{u}) - Y_d(s, t, 0)| = o_p(1).$$

In verifying (C) for these $\{\eta_i\}$, one has an analogue of (11):

$$\sup_{0\le t\le 1-\delta} [G_d(t + \delta) - G_d(t)]$$
$$\le 2k\{B \max_i \|\mathbf{c}_i\| + bn^{-1/2}\} + \sup_{0\le t\le 1-\delta} [L_d(t+\delta) - L_d(t)].$$

Note that here $G_d(t) \equiv \Sigma_i\, d_i^2 F_i(\sigma_n H^{-1}(t) + \mathbf{c}_i'\mathbf{u})$.

One similarly has an analogue of Lemma 2.3.3. Consequently, from Theorem 2.3.1 one can conclude that for each fixed $s\in[-b, b]$,

$$(31)\qquad \sup_{0\le t\le 1,\, \|\mathbf{u}\|\le B} |R_d(s, t, \mathbf{u})| = o_p(1),$$

where $R_d(s, t, u)$ equals the l.h.s. of (28) without the supremum. To complete the proof, once again exploit the compactness of $[-b, b]$ and the monotonic structure that is present in S_d and μ_d. Details are left for interested readers. □

Consider now the specialization of Theorems 2.3.1 and 2.3.2 to the case when $F_{ni} \equiv F$, F a d.f.. Note that in this case (N1) implies that $L_d(t) \equiv t$ so that (C*) is *a priori* satisfied. To state these specializations we need the following assumptions:

(F1) F *has uniformly continuous density* f *w.r.t.* λ.

(F2) $f > 0$, a.e. λ.

(F3) $\sup_{x\in\mathbb{R}} |xf(x)| \le k < \infty$.

Note that **(F1)** implies that f is bounded and that **(F2)** implies that F is strictly increasing.

Corollary 2.3.1. *Let* $X_{n1}, \ldots, X_{nn}$ *be i.i.d.* F. *In addition, suppose that* (N1), (N2), (4), (5) *and* (F1) *hold. Then* (6) *holds with* $q_{ni} = f(F^{-1})$.

If, in addition, (F2) *holds, then* (7) *holds with* $f_{ni} \equiv f$. □

Corollary 2.3.2. *Let* $X_{n1}, \ldots, X_{nn}$ *be i.i.d.* F. *In addition, suppose that* (N1), (N2), (4), (5), (F1) *and* (F3) *hold. Then* (28) *holds with* $H \equiv F$ *and* $q_{ni} \equiv f(F^{-1})$.

If, in addition, (F2) *holds, then* (29) *holds with* $f_{ni} \equiv f$. □

We shall now apply the above results to the model (1.1.1) and the $\{V_j\}$-processes of (1.1.2). The results thus obtained are useful in studying the asymptotic distributions of certain goodness-of-fit tests and a class of M-estimators of $\boldsymbol{\beta}$ of (1.1.1) when there is an unknown scale parameter also.

We need the following assumption about the design matrix **X**.

(NX) $(\mathbf{X}'\mathbf{X})^{-1}$ *exists*, $n \ge p$; $\max_i \mathbf{x}_{ni}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_{ni} = o(1)$.

This is Noether's condition for the design matrix **X**. Now, let

$$\text{(32)} \qquad \mathbf{A} = (\mathbf{X}'\mathbf{X})^{-1/2}, \qquad \mathbf{D} := \mathbf{XA},$$

$$\mathbf{q}'(t) := (q_{n1}(t), \ldots, q_{nn}(t)), \qquad \boldsymbol{\Lambda}(t) := \operatorname{diag}(\mathbf{q}(t)),$$

$$\boldsymbol{\Gamma}_1(t) := \mathbf{AX}'\boldsymbol{\Lambda}(t)\,\mathbf{XA}, \qquad \boldsymbol{\Gamma}_2(t) := n^{-1/2}H^{-1}(t)\,\mathbf{D}'\mathbf{q}(t), \qquad 0 \le t \le 1.$$

Write $\mathbf{D} = ((d_{ij}))$, $1 \le i \le n$, $1 \le j \le p$, and let $\mathbf{d}_{(j)}$ denote the j^{th} column of $\mathbf{D}$. Note that $\mathbf{D}'\mathbf{D} = \mathbf{I}_{pxp}$. This in turn implies that

(33) (N1) is satisfied by $\mathbf{d}_{(j)}$ for all $1 \le j \le p$.

Moreover, with $\mathbf{a}_{(j)}$ denoting the j^{th} column of $\mathbf{A}$,

$$(34) \quad \max_i d_{ij}^2 = \max_i (\mathbf{x}_i'\mathbf{a}_{(j)})^2 \le \max_i \sum_{j=1}^{p} (\mathbf{x}_i'\mathbf{a}_{(j)})^2$$
$$= \max_i \mathbf{x}_i' \Big(\sum_{j=1}^{p} \mathbf{a}_{(j)}\mathbf{a}_{(j)}'\Big) \mathbf{x}_i$$
$$= \max_i \mathbf{x}_i' (\mathbf{X}'\mathbf{X})^{-1} \mathbf{x}_i = o(1), \quad \text{by (NX)}.$$

Let

$$(35) \qquad L_j(t) := \sum_{i=1}^{n} d_{ij}^2 F_i(H^{-1}(t)), \qquad 0 \le t \le 1, \ 1 \le j \le p.$$

We are now ready to state

Theorem 2.3.3. *Let* $\{(\mathbf{x}_{ni}', Y_{ni}),\ 1 \le i \le n\}$, β, $\{F_{ni},\ 1 \le i \le n\}$ *be as in the model* (1.1.1). *In addition, assume that* $\{F_{ni}\}$ *satisfy* (3a), (3b) *and that* (C*) *is satisfied by each* L_j *of* (35), $1 \le j \le p$.

Then, for every $0 < B < \infty$,

$$(36) \qquad \sup \|\mathbf{A}\{\mathbf{V}(H^{-1}(t), \beta + \mathbf{A}\mathbf{u}) - \mathbf{V}(H^{-1}(t), \beta)\} - \Gamma_1(t)\mathbf{u}\| = o_p(1).$$

where the supremum is over $0 \le t \le 1$, $\mathbf{u} \in \mathcal{N}(B)$.

If, in addition, H *is strictly increasing for all* $n \ge 1$, *then, for every* $0 < B < \infty$,

$$(37) \qquad \sup \|\mathbf{A}\{\mathbf{V}(x, \beta + \mathbf{A}\mathbf{u}) - \mathbf{V}(x, \beta)\} - \Gamma_1(H(x))\mathbf{u}\| = o_p(1).$$

where the supremum is over $-\infty \le x \le \infty$, $\mathbf{u} \in \mathcal{N}(B)$.

Theroem 2.3.4. *Suppose that* $\{(\mathbf{x}_{ni}', Y_{ni}),\ 1 \le i \le n\}$ *and* $\beta \in \mathbb{R}^p$ *obey the model*

$$(38) \qquad Y_{ni} = \mathbf{x}_{ni}'\beta + \gamma\, \epsilon_{ni}, \qquad 1 \le i \le n, \ \gamma > 0,$$

with $\{\epsilon_{ni}\}$ *independent r.v.'s having d.f.'s* $\{F_{ni}\}$. *Assume that* (NX) *holds. In addition, assume that* $\{F_{ni}\}$ *satisfy* (3a), (3b), (27) *and that* (C*) *is satisfied by each* L_j *of* (35), $1 \le j \le p$.

Then for every $0 < b, B < \infty$,

(39) $\sup \|A\{V(\alpha H^{-1}(t), \beta + Au\gamma) - V(\gamma H^{-1}(t), \beta)\} - \Gamma_1(t)u - \Gamma_2(t)v\| = o_p(1),$

where $v := n^{1/2}(\alpha-\gamma)\gamma^{-1}$, $\alpha > 0$, *and the supremum is over* $0 \le t \le 1$, $u \in \mathcal{N}(B)$ *and* $|v| \le b$.

If, in addition, H *is strictly increasing for every* $n \ge 1$, *then*

(40) $\sup\|A\{V(\alpha x, \beta + Au\gamma) - V(\gamma x, \beta)\} - \Gamma_1(H(x))u - \Gamma_2(H(x))v\| = o_p(1).$

where $v := n^{1/2}(\alpha-\gamma)\gamma^{-1}$, $\alpha > 0$, *and the supremum is over* $-\infty \le x \le \infty$, $u \in \mathcal{N}(B)$ *and* $|v| \le b$.

Proof of Theorem 2.3.3. Apply Theorem 2.3.1 to $X_i = Y_i - x_i'\beta$, $c_i = x_i'A$, $1 \le i \le n$. Then F_i is the d.f. of X_i and the j^th components of $AV(H^{-1}(t), \beta + Au)$ and $AV(H^{-1}(t), \beta)$ are $S_d(t, u)$, $S_d(t, 0)$ of (1), respectively, with $d_i = d_{ij}$, $1 \le i \le n$, $1 \le j \le p$. Therefore (36) will follows by p applications of (6), one for each $d_{(j)}$, provided the assumptions of Theorem 2.3.1 are satisfied. But in view of (33) and (34), the assumption (NX) implies (N1), (N2) for $d_{(j)}$, $1 \le j \le p$. Also, (4) for the specified $\{c_i\}$ is equivalent to (NX). Finally, the C-S inequality and (33) verifies (5) in the present case. This makes Theorem 2.3.1 applicable and hence (36) follows. □

Proof of Theorem 2.3.4. Follows from Theorem 2.3.2 when applied to $X_i = (Y_i - x_i'\beta)\gamma^{-1}$, $c_i' = x_i'A$, $1 \le i \le n$, in a fashion similar to the proof of Theorem 2.3.3 above. □

The following corollaries follow from Corollaries 2.3.1 and 2.3.2 in the same way as the above Theorems 2.3.3 and 2.3.4 follow from Theorems 2.3.1 and 2.3.2. These are stated for an easy reference later on.

Corollary 2.3.3. *Suppose that the model* (1.1.1) *with* $F_{ni} \equiv F$ *holds. Assume that the design matrix* X *and the d.f.* F *satisfying* **(NX)** *and* **(F1)**. *Then,* $\forall\ 0 < B < \infty$,

(41) $\sup \|A\{V(F^{-1}(t), s) - V(F^{-1}(t), \beta)\} - f(F^{-1}(t))A^{-1}(s-\beta)\| = o_p(1).$

where the supremum is over $0 \le t \le 1$; $s \in \mathbb{R}^p$, $\|A^{-1}(s-\beta)\| \le B$.

If, in addition, F *satisfies* **(F2)**, *then*

(42) $\sup \|A\{V(x, s) - V(x, \beta)\} - f(x)\, A^{-1}(s-\beta)\| = o_p(1).$

where the supremum is over $-\infty \le x \le \infty$; $s \in \mathbb{R}^p$, $\|A^{-1}(s-\beta)\| \le B$. □

Corollary 2.3.4. *Suppose that the model* (38) *with* $F_{ni} \equiv F$ *holds and that the design matrix* X *and the d.f.* F *satisfy* (NX), (F1) *and* (F3). *Then* (39) *holds with*

$$(43) \quad \Gamma_1(t) = f(F^{-1}(t))I_{p\times p}, \qquad \Gamma_2(t) = F^{-1}(t)\, f(F^{-1}(t))AX'1, \quad 0 \le t \le 1.$$

If, in addition, F *satisfies* (F2), *then* (40) *holds with* $\Gamma_j(H) \equiv \Gamma_j(F)$, j = 1, 2. I.e.,

$$(44) \qquad \sup \|A\{V(\alpha x, \beta + Au\gamma) - V(\gamma x, \beta)\} - f(x)u - xf(x)v\| = o_p(1),$$

where the supremum is over $-\infty \le x \le \infty$; $u \in \mathcal{N}(B)$ *and* $|v| \le b$, *with* v *as in* (39). □

We end this section by stating an a.u.l. result about the ordinary residual empirical processes H_n of (1.2.1) for an easy reference later on.

Corollary 2.3.5. *Suppose that the model* (1.1.1) *with* $F_{ni} \equiv F$ *holds. Assume that the design matrix* X *and the d.f.* F *satisfying* (NX) *and* (F1). *Then,* $\forall\ 0 < B < \infty$,

$$(45) \quad \sup|n^{1/2}\{H_n(F^{-1}(t), s) - H_n(F^{-1}(t), \beta)\} - \\ - f(F^{-1}(t))\cdot n^{-1/2}\Sigma_i\, x_{ni}'\, A\cdot A^{-1}(s - \beta)| = o_p(1),$$

where the supremum is over $0 \le t \le 1$; $s \in \mathbb{R}^p$, $\|A^{-1}(s - \beta)\| \le B$.

If, in addition, F *satisfies* (F2), *then,* $\forall\ 0 < B < \infty$,

$$(46) \quad \sup|n^{1/2}\{H_n(x, s) - H_n(x, \beta)\} - f(x)\cdot n^{-1/2}\Sigma_i\, x_{ni}'A\cdot A^{-1}(s-\beta)| = o_p(1).$$

where the supremum is over $-\infty \le x \le \infty$; $s \in \mathbb{R}^p$, $\|A^{-1}(s - \beta)\| \le B$.

Proof. The proof follows from Theorem 2.3.1 by specializing it to the case where $d_{ni} \equiv n^{-1/2}$ and the rest of the entities as in the proof of Theorem 2.3.3. □

Note: Ghosh and Sen (1971) and Koul and Zhu (1991) have proved an almost sure version of (42) in the case p = 1 and p > 1, respectively. □

2.4. SOME FURTHER PROBABILISTIC RESULTS FOR W.E.P.'S.

For the sake of general interest, here we state some further results about w.e.p.'s. To begin with, we have

2.4.1. Laws of the iterated logarithm:

In this subsection, we assume that

(1) $$d_{ni} \equiv d_i, \quad \eta_{ni} \equiv \eta_i, \quad G_{ni} \equiv G_i, \quad 1 \le i \le n.$$

Define

(2) $$\mathcal{U}_n(t) := \sum_{i=1}^{n} d_i \{I(\eta_i \le t) - G_i(t)\}, \qquad \sigma_n^2 := \sum_{i=1}^{n} d_i^2,$$

$$\xi_n(t) := \mathcal{U}_n(t)/\{2\sigma_n^2 \, \ell n\, \ell n\, \sigma_n^2\}^{1/2}, \qquad n \ge 1, \ 0 \le t \le 1.$$

Let $r(s,t) := s \wedge t - st$, $0 \le s, t \le 1$, and $H(r)$ be the reproducing kernel Hilbert space generated by the kernel r with $\|\cdot\|_r$ denoting the associated norm on $H(r)$. Let

(3) $$K = \{f \in H(r); \ \|f\|_r \le 1\}.$$

Theorem 2.4.1. *If* $\eta_1, \eta_2, \ldots$ *are i.i.d. uniform on* $[0, 1]$ *and* $d_1, d_2, \ldots$ *are any real numbers satisfying*

(a) $$\lim_n \sigma_n^2 = \infty, \qquad \lim_n \left(\max_{1 \le i \le n} d_i^2\right) \frac{\ell n\ \ell n\ \sigma_n^2}{\sigma_n^2} = 0,$$

then

$$P(\mathcal{d}(\xi_n, K) \longrightarrow 0 \text{ and the set of limit points of } \{\xi_n\} \text{ is } K) = 1. \qquad \square$$

Theorem 2.4.1 was proved by Vanderzanden (1980, 1984) using some of the results of Kuelbs (1976) and certain martingale properties of ξ_n.

Theorem 2.4.2. *Let* $\eta_1, \eta_2, \ldots.$ *be independent nonnegative r.v.'s. Let* $\{d_i\}$ *be any real numbers. Then*

$$\limsup_n \sup_{t \ge 0} \sigma_n^{-1} |\, \mathcal{U}_n(t-)| < \infty, \text{ a.s..} \qquad \square$$

A proof of this appears in Marcus and Zinn (1984). Actually they prove some other interesting results about w.e.p.'s with weights which are r.v.'s and functions of t. Most of their results, however, are concerned with the bounded law of the iterated logarithm. They also proved the following inequality that is similar to, yet a generalization of, the classical Dvoretzky-Kiefer-Wolfowitz exponential inequality for the ordinary empirical process. Their proof is valid for triangular arrays and real r.v.'s.

Exponential inequality. *Let* $X_{n1}, X_{n2}, \ldots, X_{nn}$ *be independent r.v.'s with respective d.f.'s* $F_{n1}, \ldots, F_{nn}$ *and* $\{d_{ni}\}$ *be any real numbers satisfying* (N1). *Then,* $\forall\ \lambda > 0, \forall\ n \ge 1,$

(4) $P(\sup_{|x|<\infty} |\sum_{i=1}^{n} d_{ni}\{I(X_{ni} \le x) - F_{ni}(x)\}| \ge \lambda) \le [1+(8\pi)^{1/2}\lambda]\exp(-\lambda^2/8)$. □

The above two theorems immediately suggest some interesting probabilistic questions. For example, is Vanderzanden's result valid for nonidentical r.v.'s $\{\eta_i\}$? Or can one remove the assumption of nonnegative $\{\eta_i\}$ in Theorem 2.4.1?

2.4.2. Weak convergence of w.e.p.'s in $\mathbb{D}[0, 1]^p$, in $\mathscr{d}_q$ -metric and an embedding result.

Next, we state a weak convergence result for multivariate r.v.'s. For this we revert back to triangular arrays. Now suppose that $\boldsymbol{\eta}_{ni} \in [0, 1]^p$, $1 \le i \le n$, are independent r.v.'s of dimension p. Define

$$W_d(\mathbf{t}) := \sum_{i=1}^{n} d_{ni}\{I(\boldsymbol{\eta}_{ni} \le \mathbf{t}) - G_{ni}(\mathbf{t})\}, \quad \mathbf{t} \in [0, 1]^p. \tag{5}$$

Let G_{nij} be the j^{th} marginal of G_{ni}, $1 \le i \le n$, $1 \le j \le p$.

Theorem 2.4.3. *Let* $\{\boldsymbol{\eta}_{ni}, 1 \le i \le n\}$ *be independent* p*-variate r.v.'s and* $\{d_{ni}\}$ *satisfy* (N1) *and* (N2). *Moreover suppose that for each* $1 \le j \le p$,

$$\lim_{\delta\to 0} \limsup_n \sup_{0\le t\le 1-\delta} \sum_{i=1}^{n} d_{ni}^2\{G_{nij}(t+\delta) - G_{nij}(t)\} = 0.$$

Then, for every $\epsilon > 0$

(i) $$\lim_{\delta\to 0} \limsup_n P(\sup_{|\mathbf{s}-\mathbf{t}|<\delta} |W_d(\mathbf{t}) - W_d(\mathbf{s})| > \epsilon) = 0.$$

(ii) *Moreover,* $W_d \Rightarrow$ *some* W *on* $(\mathbb{D}[0,1]^p, \mathscr{d})$ *if, and only if, for each* $\mathbf{s}, \mathbf{t} \in [0, 1]^p$, $\mathrm{Cov}(W_d(\mathbf{s}), W_d(\mathbf{t})) \longrightarrow \mathrm{Cov}(W(\mathbf{s}), W(\mathbf{t})) =: \mathcal{C}(\mathbf{s}, \mathbf{t})$.
In this case W *is necessarily a Gaussian process,* $P(W \in \mathbb{C}[0, 1]^p) = 1$, $W(\mathbf{0}) = 0 = W(\mathbf{1})$. □

Theorem 2.4.3 is essentially proved in Vanderzanden (1980), using results of Bickel and Wichura (1971).

Mehra and Rao (1975), Withers (1975), and Koul (1977), among others, obtain the weak convergence results for $\{W_d\}$-processes when $\{\eta_{ni}\}$ are weakly dependent. See Dehling and Taqqu (1989) and Koul and Mukherjee (1992) for similar results when $\{\eta_{ni}\}$ are long range dependent.

Shorack (1979) proved the weak convergence of W_d/q-process in the $\mathscr{d}$-metric, where $q \in \mathcal{Q}$, with

$$\mathcal{Q} := \{q;\ q \text{ a continuous function on } [0, 1],\ q \geq 0,\ q(t) = q(1-t),\ q(t) \uparrow \text{ and } t^{-1/2}q(t) \downarrow \text{ for } 0 \leq t \leq 1/2,\ \int_0^1 q^{-2}(t)dt < \infty\}.$$

Theorem 2.4.4. *Suppose that* $\eta_{n1}, \ldots, \eta_{nn}$ *are independent r.v.'s in* $[0, 1]$ *with respective d.f.'s* $G_{n1}, \ldots, G_{nn}$ *such that*

$$n^{-1}\sum_{i=1}^{n} G_{ni}(t) = t, \quad 0 \leq t \leq 1.$$

In addition, suppose that $\{d_{ni}\}$ *satisfy* (**N1**) *and* (**B**). *Then,*

(i) $\forall\ \epsilon > 0,\ \forall\ q \in \mathcal{Q}$,

$$\lim_{\delta \to 0} \limsup_n P\Big(\sup_{|t-s|<\delta} \Big|\frac{W_d(t)}{q(t)} - \frac{W_d(s)}{q(s)}\Big| > \epsilon\Big) = 0.$$

(ii) $q^{-1}W_d \Rightarrow q^{-1}W$, W *a continuous Gaussian process with covariance function* C *if, and only if* $C_d \longrightarrow C$. □

Shorack (1991) and Einmahl and Mason (1991) proved the following embedding result.

Theorem 2.4.5. *Suppose that* $\eta_{n1}, \ldots, \eta_{nn}$ *are i.i.d. Uniform* $[0, 1]$ *r.v.'s. In addition, suppose that* $\{d_{ni}\}$ *satisfy* (**N1**) *and that*

$$\sum_{i=1}^{n} d_{ni} = 0, \qquad n\sum_{i=1}^{n} d_{ni}^4 = O(1).$$

Then on a rich enough probability space there exist a sequence of versions $\mathcal{W}_d$ *of the processes* W_d *and a fixed Brownian bridge* B *on* $[0, 1]$ *such that*

$$\sup_{1/n \leq t \leq 1-1/n} n^{\nu} \frac{|\mathcal{W}_d(t) - B(t)|}{\{t(1-t)\}^{1/2-\nu}} = O_p(1), \qquad \textit{for all } 0 \leq \nu < 1.$$

The closed interval $1/n \leq t \leq 1-1/n$ *may be replaced by the open interval* $\min\{\eta_{nj};\ 1 \leq j \leq n\} < t < \max\{\eta_{nj};\ 1 \leq j \leq n\}$. □

2.4.3. A martingale property.

In this subsection we shall prove a martingale property of w.e.p.'s. Let $X_{n1}, X_{n2}, \ldots, X_{nn}$ be independent real r.v.'s with respective d.f.'s $F_{n1}, \ldots, F_{nn}$; $d_{n1}, \ldots, d_{nn}$ be real numbers. Let $a \leq b$ be fixed real numbers. Define,

$$M_n(t) := \sum_{i=1}^{n} d_{ni}\{I(X_{ni} \in (a, t] - p_{ni}(a, t]\}\ \{1 - p_{ni}(a, t]\}^{-1},$$

$$R_n(t) := \sum_{i=1}^{n} d_{ni}\{I(X_{ni} \in (t, b] - p_{ni}(t, b]\} \{1 - p_{ni}(t, b]\}^{-1}, \qquad t \in \mathbb{R},$$

where

$$p_{ni}(s,t] := F_{ni}(t) - F_{ni}(s), \quad 0 \leq s \leq t \leq 1, \quad 1 \leq i \leq n.$$

Let $T_1 \subset [a, \infty)$, $T_2 \subset (-\infty, b]$ be such that $M_n(t)$ $[R_n(t)]$ is well-defined for $t \in T_1$ $[t \in T_2]$. Let

$$\mathcal{F}_{1n}(t) := \sigma\text{-field } \{I(X_{ni} \in (a, s]), \quad a \leq s \leq t, \quad i = 1, ..., n\}, \quad t \in T_1,$$

$$\mathcal{F}_{2n}(t) := \sigma\text{-field } \{I(X_{ni} \in (s, b]), \quad t \leq s \leq b, \quad i = 1, ..., n\}, \quad t \in T_2.$$

Martingale Lemma. *Under the above set up, for each* $n \geq 1$, $\{M_n(t), \mathcal{F}_{1n}(t), t \in T_1\}$ *is a martingale and* $\{R_n(t), \mathcal{F}_{2n}(t), t \in T_2\}$ *is a reverse martingale.*

Proof. Write $q_i(a, s] = 1 - p_i(a, s]$. Because $\{X_i\}$ are independent, for $a \leq s \leq t$

$$\begin{aligned}
&E\{M_n(t) | \mathcal{F}_{1n}(s)\} \\
&= \sum_{i=1}^{n} d_i\{q_i(a,t]\}^{-1}[I(X_i \in(a, s])\, E\{(I(X_i \in(a, t]) - p_i(a, t]) \mid X_i \in(a, s]\} \\
&\qquad + I(X_i \notin(a, s])\, E\{(I(X_i \in(a, t]) - p[_i(a, t]) \mid X_i \notin(a, s]\}] \\
&= \sum_{i=1}^{n} d_i \{q_i(a, t)\}^{-1}[I(X_i \in(a, s]q_i(a, t] + \\
&\qquad + I(X_i \notin(a, s] \{\frac{p_i(s, t]}{q_i(a, s]} - p_i(a, t]\}] \\
&= \sum_{i=1}^{n} d_i\{q_i(a, s)\}^{-1}(I(X_i \in(a, s]) - q_i(a, s]\} = M_n(s).
\end{aligned}$$

A similar argument yields the result about R_n. □

Note. In the case $\{X_{ni}\}$ are i.i.d. and $d_{ni} \equiv n^{-1/2}$, this Lemma is well known. In the case $\{X_{ni}\}$ are i.i.d. and $\{d_{ni}\}$ are arbitrary, the observation about $\{M_n\}$ being a martingale first appeared in Sinha and Sen (1979). The above Martingale Lemma appears in Vanderzanden (1980, 1984).

Theorem 2.4.1 above generalizes a result of Finkelstein (1971) for the ordinary empirical process to w.e.p.'s of i.i.d. r.v.'s.. In fact, the set K is the same as the set K of Finkelstein. □□

CHAPTER 3

LINEAR RANK AND SIGNED RANK STATISTICS

3.1. INTRODUCTION

Let $\{X_{ni}, F_{ni}\}$ be as in (2.2.33) and $\{\mathbf{c}_{ni}\}$ be p×1 real vectors. The rank and the absolute rank of the ith residual are defined, respectively, as

$$(1)\quad R_{i\mathbf{u}} = \sum_{j=1}^{n} I(X_{nj} - \mathbf{u}'\mathbf{c}_{nj} \le X_{ni} - \mathbf{u}'\mathbf{c}_{ni}),$$

$$R^{+}_{i\mathbf{u}} = \sum_{j=1}^{n} I(|X_{nj} - \mathbf{u}'\mathbf{c}_{nj}| \le |X_{ni} - \mathbf{u}'\mathbf{c}_{ni}|), \qquad 1 \le i \le n, \ \ \mathbf{u} \in \mathbb{R}^p.$$

Let φ be a nondecreasing real valued function on $[0, 1]$ and define

$$(2)\quad T_d(\varphi, \mathbf{u}) = \sum_{i=1}^{n} d_{ni}\, \varphi(R_{i\mathbf{u}}/(n+1)),$$

$$T^{+}_d(\varphi, \mathbf{u}) = \sum_{i=1}^{n} d_{ni}\, \varphi^{+}(R^{+}_{i\mathbf{u}}/(n+1))\, s(X_{ni} - \mathbf{u}'\mathbf{c}_{ni}), \qquad \mathbf{u} \in \mathbb{R}^p,$$

where $\varphi^{+}(s) = \varphi((s+1)/2)$, $0 \le s \le 1$, and $s(x) = I(x > 0) - I(x < 0)$.

The processes $\{T_d(\varphi, \mathbf{u}),\ \mathbf{u} \in \mathbb{R}^p\}$ and $\{T^{+}_d(\varphi, \mathbf{u}),\ \mathbf{u} \in \mathbb{R}^p\}$ are used to define rank (R) estimators of $\boldsymbol{\beta}$ in the linear regression model (1.1.1). See, e.g., Adichie (1967), Koul (1971), Jurečková (1971) and Jaeckel (1972). One key property used in studying these R-estimators is the asymptotic uniform linearity (a.u.l.) of $T_d(\varphi, \mathbf{u})$ and $T^{+}_d(\varphi, \mathbf{u})$ in $\mathbf{u} \in \mathscr{N}(B)$. Such results have been proved by Jurečková (1969) for $T_d(\varphi, \mathbf{u})$ for general but fixed functions φ, by Koul (1969) for $T^{+}_d(I, \mathbf{u})$ (where I is the identity function) and by Van Eeden (1971) for $T^{+}_d(\varphi, \mathbf{u})$ for general but fixed φ functions. In all of these papers $\{X_{ni}\}$ are assumed to be i.i.d..

In Sections 3.2 and 3.3 below we prove the a.u.l. of $T_d(\varphi, .)$, $T^{+}_d(\varphi, .)$, uniformly in those φ which have $\|\varphi\|_{tv} < \infty$, and under fairly general independent setting. These proofs reveal that this a.u.l. property is also a consequence of the asymptotic continuity of certain w.e.p.'s and the smoothness of $\{F_{ni}\}$.

Besides being useful in studying the asymptotic distributions of R-estimators of $\boldsymbol{\beta}$ these results are also useful in studying some rank based

minimum distance estimators, some goodness-of-fit tests for the error distributions of (1.1.1) and the robustness of R-estimators against certain heteroscedastic errors.

3.2. ASYMPTOTIC UNIFORM LINEARITY OF LINEAR RANK STATISTICS

At the outset we shall assume

(1) $\varphi \in \mathscr{C} := \{\varphi\colon [0,1] \longrightarrow \mathbb{R},\ \varphi \in \mathbb{D}\mathcal{I}[0,1], \text{ with } \|\varphi\|_{tv} := \varphi(1) - \varphi(0) = 1\}$.

Define the w.e.p. based on ranks, with weights $\{d_{ni}\}$,

(2) $$Z_d(t, \mathbf{u}) := \Sigma_i\, d_{ni}\, I(R_{i\mathbf{u}} \le nt), \qquad 0 \le t \le 1,\ \mathbf{u} \in \mathbb{R}^p.$$

Note that

(3) $$\begin{aligned} T_d(\varphi, \mathbf{u}) &= \int \varphi(nt/(n+1))\, Z_d(dt,\mathbf{u}) \\ &= -\int Z_d((n+1)t/n, \mathbf{u})\, d\varphi(t) + n\bar{d}_n\, \varphi(1), \qquad n\bar{d}_n = \sum_{i=1}^{n} d_{ni}. \end{aligned}$$

The representation (3) shows that in order to prove the a.u.l. of $T_d(\varphi, .)$, it suffices to prove it for $Z_d(t, .)$, uniformly in $0 \le t \le 1$. Thus, we shall first prove the a.u.l. property for the Z_d-process. Define, for $x \in \mathbb{R}$, $0 \le t \le 1$, $\mathbf{u} \in \mathbb{R}^p$,

(4) $$\begin{aligned} H_{n\mathbf{u}}(x) &= n^{-1}\, \Sigma_i\, I(X_{ni} - \mathbf{c}_{ni}'\mathbf{u} \le x), & H_{\mathbf{u}}(x) &:= n^{-1}\Sigma_i\, F_{ni}(x + \mathbf{c}_{ni}'\mathbf{u}), \\ H_{n\mathbf{u}}^{-1}(t) &= \inf\{x;\ H_{n\mathbf{u}}(x) \ge t\}, & H_{\mathbf{u}}^{-1}(t) &= \inf\{x;\ H_{\mathbf{u}}(x) \ge t\}. \end{aligned}$$

Note that H_0 is the H of (2.2a.33). We shall write H_n for H_{n0}.

Recall that for any d.f. G,

$$G(G^{-1}(t)) \ge t,\quad 0 \le t \le 1 \quad \text{and} \quad G^{-1}(G(x)) \le x, \quad x \in \mathbb{R}.$$

This fact and the relation $nH_{n\mathbf{u}}(X_i - \mathbf{c}_i'\mathbf{u}) \equiv R_{i\mathbf{u}}$ yield that $\forall\ 0 \le t \le 1$,

(5) $$[X_i - \mathbf{c}_i'\mathbf{u} \ge H_{n\mathbf{u}}^{-1}(t)] \Rightarrow [R_{i\mathbf{u}} \ge nt] \Rightarrow [X_i - \mathbf{c}_i'\mathbf{u} \ge H_{n\mathbf{u}}^{-1}(t)], \quad 1 \le i \le n.$$

For technical convenience, it is desirable to center the weights of linear rank statistics appropriately. Accordingly, let

(6) $$w_{ni} := (d_{ni} - \bar{d}_n), \qquad 1 \le i \le n.$$

Then, with Z_w denoting the Z_d when weights are $\{w_{ni}\}$,

$$Z_d(t, u) = Z_w(t, u) + \bar{d}_n \cdot [nt], \qquad 0 \le t \le 1, u \in \mathbb{R}^p.$$

Hence

$$(7) \qquad Z_d(t, u) - Z_d(t, 0) = Z_w(t, u) - Z_w(t, 0), \qquad 0 \le t \le 1, u \in \mathbb{R}^p.$$

Next define, for arbitrary real weights $\{d_{ni}\}$,

$$(8) \qquad \mathcal{Y}_d(t, u) := \Sigma\, d_{ni}\, I(X_{ni} - c_{ni}' u \le H_{nu}^{-1}(t)), \qquad 0 \le t \le 1, u \in \mathbb{R}^p.$$

By (5) and direct algebra, for any weights $\{d_{ni}\}$,

$$(9) \qquad \sup_{t, u} |Z_d(t, u) - \mathcal{Y}_d(t, u)| \le 2 \max_i |d_i|.$$

Consider the condition

$$(\mathbf{N3}) \qquad \tau_w^2 = 1, \qquad \max_i w_{ni}^2 \longrightarrow 0.$$

In view of (7) and (9), (N3) implies that the problem of proving the a.u.l. for the Z_d-process is reduced to proving it for the $\mathcal{Y}_w$-process.
Recall the definitions in (2.3.1) and define

$$(10) \qquad \tilde{T}_d(t, u) := \mathcal{Y}_d(t, u) - \mu_d(t, u), \qquad 0 \le t \le 1, u \in \mathbb{R}^p.$$

Note the basic decomposition: for any real numbers $\{d_{ni}\}$ and for all $0 \le t \le 1$, $u \in \mathbb{R}^p$,

$$(11) \qquad \tilde{T}_d(t, u) = Y_d(HH_{nu}^{-1}(t), u) + \mu_d(HH_{nu}^{-1}(t), u) - \mu_d(t, u),$$

provided H is strictly increasing for all $n \ge 1$. Decomposition (11) is basic to the following proof of the a.u.l. property of Z_d.

Theorem 3.2.1. *Suppose that* $\{X_{ni}, F_{ni}\}$ *satisfy* (2.2a.34), (N3) *holds, and* $\{c_{ni}\}$ *satisfy* (2.3.4) *and* (2.3.5) *with* $d_{ni} \equiv w_{ni}$. *In addition, assume that* (C*) *holds with* $d_{ni} \equiv w_{ni}$, H *is strictly increasing, the densities* $\{f_{ni}\}$ *of* $\{F_{ni}\}$ *satisfy* (2.3.3b), *and that*

$$(12) \qquad \lim_{\delta \to 0} \limsup_n \max_i \sup_{|H(x) - H(y)| < \delta} |f_{ni}(x) - f_{ni}(y)| = 0.$$

Then, for every $0 < B < \infty$,

$$(13) \qquad \sup |\tilde{T}_w(t, u) - Y_w(t, 0) - \mu_w(HH_{nu}^{-1}(t), 0) + \mu_w(t, 0)| = o_p(1)$$

where the supremum is being taken over $0 \le t \le 1$, $\mathbf{u} \in \mathbb{R}^p$.

Before proceeding to prove the theorem, we prove the following lemma which is of independent interest. In this result, no assumptions other than independence of $\{X_{ni}\}$ are being used.

Lemma 3.2.1. *Let* H, H_n, $H_\mathbf{u}$ *and* $H_{n\mathbf{u}}$ *be as in* (4) *above. Assuming only* (2.2a.34), *we have*

$$\|H_n - H\|_\infty \longrightarrow 0 \quad \text{a.s..} \tag{14}$$

If, in addition, (2.3.4) *holds and if, for any* $0 < B < \infty$,

$$\sup_{|x-y| \le 2m_n B} |H(x) - H(y)| \longrightarrow 0, \qquad (m_n = \max_i \|\mathbf{c}_i\|), \tag{15}$$

then,

$$\sup_{|x| < \infty,\, \|\mathbf{u}\| \le B} |H_{n\mathbf{u}}(x) - H_\mathbf{u}(x)| \longrightarrow 0 \quad \text{a.s..} \tag{16}$$

Proof. Note that $H_n(x) - H(x)$ is a sum of centered independent Bernoulli r.v.'s. Thus $E[H_n(x) - H(x)]^4 = O(n^{-2})$. Apply the Markov inequality with the 4th moment and the Borel-Cantelli lemma to obtain

$$|H_n(x) - H(x)| \longrightarrow 0, \text{ a.s.,} \quad \text{for every } x \in \mathbb{R}.$$

Now proceed as in the proof of the Glivenko-Cantelli Lemma (Loève (1963), p.21) to conclude (14).

To prove (16), note that $\mathbf{u} \in \mathcal{N}(B)$ implies that $-m_n B \le \mathbf{c}_i' \mathbf{u} \le m_n B$, $1 \le i \le n$. The monotonicity of $H_{n\mathbf{u}}$ and $H_\mathbf{u}$ yields that for $\mathbf{u} \in \mathcal{N}(B)$, $x \in \mathbb{R}$,

$$\begin{aligned} &H_n(x-Bm_n) - H(x-Bm_n) + H(x-Bm_n) - H(x+Bm_n) \\ &\quad \le H_{n\mathbf{u}}(x) - H_\mathbf{u}(x) \\ &\qquad \le H_n(x+Bm_n) - H(x+Bm_n) + H(x+Bm_n) - H(x-Bm_n). \end{aligned}$$

Hence (16) follows from (15) and the following inequality:

$$\text{l.h.s. } (16) \le 2 \sup_{|x| < \infty} |H_n(x) - H(x)| + \sup_{|x-y| \le 2m_n B} |H(x) - H(y)|. \qquad \square$$

Proof of Theorem 3.2.1. From (11), for all $0 \le t \le 1$, $\mathbf{u} \in \mathbb{R}^p$,

$$\tilde{T}_w(t,\mathbf{u}) = [Y_w(HH_{n\mathbf{u}}^{-1}(t), \mathbf{u}) - Y_w(HH_{n\mathbf{u}}^{-1}(t), 0]$$

$$+ [Y_w(HH_{n\mathbf{u}}^{-1}(t), 0) - Y_w(t,0)]$$

$$+ Y_w(t, 0) - [\mu_w(t, \mathbf{u}) - \mu_w(t, 0) - \mathbf{u}'\nu_w(t)]$$

$$+ [\mu_w(HH_{n\mathbf{u}}^{-1}(t), \mathbf{u}) - \mu_w(HH_{n\mathbf{u}}^{-1}(t), 0) - \mathbf{u}'\nu_w(HH_{n\mathbf{u}}^{-1}(t))]$$

$$+ \mu_w(HH_{n\mathbf{u}}^{-1}(t), 0) - \mu_w(t, 0) + \mathbf{u}'[\nu_w(HH_{n\mathbf{u}}^{-1}(t)) - \nu_w(t)].$$

Therefore,

$$\text{l.h.s. } (13) \le \sup|Y_w(t, \mathbf{u}) - Y_w(t, 0)| + \sup|Y_w(HH_{n\mathbf{u}}^{-1}(t), 0) - Y(t,0)|$$

$$+ 2 \sup |\mu_w(t, \mathbf{u}) - \mu_w(t, 0) - \mathbf{u}'\nu_w(t)|$$

$$+ \sup |\mathbf{u}'[\nu_w(HH_{n\mathbf{u}}^{-1}(t)) - \nu_w(t)]|$$

$$= A_1 + A_2 + A_3 + A_4, \quad \text{say,} \tag{17}$$

where, as usual, the supremum is being taken over $0 \le t \le 1$, $\mathbf{u} \in \mathcal{N}(B)$. In what follows, the range of x and y over which the supremum is being taken is $\mathbb{R}$, unless specified otherwise.

Now, (2.3.3b) implies that $|H(x) - H(y)| \le |x-y|\, k$. This and (2.3.4) together imply (15). It also implies that

$$\sup_{|x-y|<\delta} |f_{ni}(y) - f_{ni}(x)| \le \sup_{|H(x)-H(y)|<k\delta} |f_{ni}(y) - f_{ni}(x)|.$$

for all $1 \le i \le n$ and all $\delta > 0$. Hence, by (12), it follows that $\{f_{ni}\}$ satisfy (2.3.3a). Now apply Lemma 2.3.1 and (2.3.25), with $d_{ni} = w_{ni}$, $1 \le i \le n$, to conclude that

$$A_j = o_p(1), \quad j = 1, 3. \tag{18}$$

Next, observe that

$$\sup|HH_{n\mathbf{u}}^{-1}(t) - t| \le \sup_{x,\mathbf{u}} |H_{n\mathbf{u}}(x) - H_{\mathbf{u}}(x)| + \sup_{x,\mathbf{u}} |H_{\mathbf{u}}(x) - H(x)| + n^{-1}, \tag{19}$$

$$\sup_{x,\mathbf{u}}|H_{\mathbf{u}}(x) - H(x)| \le \sup_{x}|H(x + m_n B) - H(x - m_n B)|.$$

Hence, in view of (19) and Lemma 3.2.1, we obtain

(20) $$\sup_{t,\mathbf{u}} |HH_{n\mathbf{u}}^{-1}(t) - t| \longrightarrow 0, \text{ a.s..}$$

(We need to use the convergence in probability only).

Now, fix a $\delta > 0$ and let $B_n^\delta = [\sup_{t,\mathbf{u}} | HH_{n\mathbf{u}}^{-1}(t) - t| < \delta]$. By (20),

(21) $$\limsup_n P((B_n^\delta)^c) = 0.$$

Now observe that $Y_d(.,\mathbf{0}) = W_d^*(.)$ of (2.2a.33). Hence, with A_2 as in (17), for every $\eta > 0$,

(22) $$\limsup_n P(|A_2| \geq \eta) \leq \limsup_n P(\sup_{|t-s|<\delta} |W_w^*(t) - W_w^*(s)| \geq \eta, B_n^\delta).$$

Upon letting $\delta \longrightarrow 0$ in (22), (2.2a.35) implies

(23) $$A_2 = o_p(1).$$

Next, we have

(24) $$\begin{aligned} &\lim_{\delta\to0} \limsup_n \sup_{|t-s|<\delta} \|\boldsymbol{\nu}_w(t) - \boldsymbol{\nu}_w(s)\| \\ &\leq \lim_{\delta\to0} \limsup_n \max_i \sup_{|H(x)-H(y)|<\delta} |f_{ni}(y) - f_{ni}(x)|(\Sigma_i \|\mathbf{w}_i\mathbf{c}_i\|) \\ &= 0, \qquad\qquad \text{by (12) and (2.3.5).} \end{aligned}$$

From (24) and (21) one obtains, in a fashion similar to (23), that

(25) $$A_4 = o_p(1).$$

This completes the proof of the theorem. □

From a practical point of view, it is worthwhile to state the a.u.l. result in the i.i.d. case separately. Accordingly, we have

Theorem 3.2.2. *Suppose that* $X_{n1}, \dots, X_{nn}$ *are i.i.d.* F. *In addition, assume that* (**F1**), (**F2**), (**N3**), (2.3.4) *and* (2.3.5) *with* $d_{ni} \equiv w_{ni}$ *hold. Then,* $\forall\ 0 < B < \infty$,

(26) $$\sup_{0\leq t\leq 1, \|\mathbf{u}\|\leq B} |Z_d(t,\mathbf{u}) - Z_d(t,\mathbf{0}) - \mathbf{u}'\Sigma_i w_{ni}\mathbf{c}_{ni}\, q(t)| = o_p(1),$$

$$(27) \qquad \sup_{\varphi \in \mathscr{C}, \|\mathbf{u}\| \le B} |T_d(\varphi, \mathbf{u}) - T_d(\varphi, 0) + \mathbf{u}' \Sigma_i w_{ni} \mathbf{c}_{ni} \int q \, d\varphi| = o_p(1).$$

where $q = f(F^{-1})$.

Proof. Let $\rho = \Sigma\, w_{ni}\, \mathbf{c}_{ni}$. From (7),

$$(28) \qquad \text{l.h.s. } (26) = \sup_{t, \mathbf{u}} |Z_w(t, \mathbf{u}) - Z_w(t, 0) - \mathbf{u}' \rho\, q(t)|.$$

Take $F_{ni} \equiv F$ in Theorem 3.2.1. Then **(F1)** and **(F2)** imply that q is uniformly continuous on $[0, 1]$ and ensure the satisfaction of all assumptions pertaining to F in Theorem 3.2.1. In addition, $\mu_w(t, 0) = 0$, $0 \le t \le 1$. Thus, Theorem 3.2.1 is applicable and one obtains

$$\sup_{t, \mathbf{u}} |\tilde{T}_w(t, \mathbf{u}) - Y_w(t, 0)| = o_p(1)$$

which in turn yields

$$(29) \qquad \sup_{t, \mathbf{u}} |\tilde{T}_w(t, \mathbf{u}) - \tilde{T}_w(t, 0)| = o_p(1).$$

From (10) and (28),

$$\begin{aligned} \text{l.h.s. } (26) &\le \sup_{t, \mathbf{u}} \{ |Z_w(t, \mathbf{u}) - \mathscr{Y}_w(t, \mathbf{u})| + |Z_w(t, 0) - \mathscr{Y}_w(t, 0)| + \\ &\qquad + |\tilde{T}_w(t, \mathbf{u}) - \tilde{T}_w(t, 0)| + |\mu_w(t, \mathbf{u}) - \mathbf{u}' \rho\, q(t)| \} \\ &= o_p(1), \end{aligned}$$

by (9), (10), **(N3)**, (29) and Lemma 2.3.1 applied to $F_{ni} \equiv F$, $d_{ni} \equiv w_{ni}$.

To conclude (27), observe that

$$\begin{aligned} \text{l.h.s.}(27) &\le \sup_{t, \mathbf{u}} \{ |Z_d(t, \mathbf{u}) - Z_d(t, 0) - \mathbf{u}' \rho\, q(t)| \\ &\qquad + |\mathbf{u}' \rho| \; |q((n+1)t/n) - q(t)| \} \\ &= o_p(1), \end{aligned}$$

by (26), the uniform continuity of q and (2.3.5) with $d_{ni} \equiv w_{ni}$. □

Remark 3.2.1. Theorem 3.2.2 continues to hold if F depends on n, provided now that the $\{q\}$ are uniformly equicontinuous on $[0, 1]$. □

Remark 3.2.2. An analogue of Theorem 3.2.2 was first proved in Koul (1970) under somewhat stronger conditions on various underlying entities. In Jurečková (1969) one finds yet another variant of (27) for a fixed but a fairly general function φ and with p in c_{ni} equal to 1. Because of the importance of the a.u.l. property of $T_d(\varphi, .)$, it is worthwhile to compare Theorem 3.2.2 above with that of Jurečková's Theorem 3.1 (1969). For the sake of completeness we state it as

Theorem 3.2.3. (Theorem 3.1, Jurečková (1969)). *Let* $X_{n1}, \ldots, X_{nn}$ *be i.i.d.* F. *In addition, assume the following:*

(a) *F has an absolutely continuous density* f *whose a.e. derivative* $\dot{f}$ *satisfies*

$$0 < I(f) < \infty, \qquad I(f) := \int (\dot{f}/f)^2 dF.$$

(b) $\{w_{ni}\}$ *satisfy* (N3).

(c) 1. $\quad \Sigma(c_{ni} - \bar{c}_n)^2 \leq M < \infty \qquad$ (*recall here* c_{ni} *is* 1×1)

2. $\quad \max(c_{ni} - \bar{c}_n)^2 = o(1), \qquad \bar{c}_n = n^{-1} \sum_{i=1}^{n} c_{ni}.$

(d) φ *is a nondecreasing function on* (0, 1) *with*

$$\int_0^1 (\varphi(t) - \bar{\varphi})^2 \, dt > 0, \quad \bar{\varphi} := \int_0^1 \varphi(u) du.$$

(e) *Either* $(d_{ni} - d_{nj})(c_{ni} - c_{nj}) \geq 0, \ \forall \ 1 \leq i, j \leq n,$

or $\quad (d_{ni} - d_{nj})(c_{ni} - c_{nj}) \leq 0, \ \forall \ 1 \leq i, j \leq n.$

Then, $\forall \ 0 < B < \infty$,

(f) $$\sup_{\|u\| \leq B} |T_d(\varphi, u) - T_d(\varphi, 0) + u \, \Sigma_i \, w_{ni} \, c_{ni} \, b(\varphi, f)| = o_p(1)$$

where $b(\varphi, f) := -\int_{-\infty}^{\infty} \varphi(F(x)) \, \dot{f}(x) \, dx.$ □

The strongest point of Theorem 3.2.3 is that it allows for unbounded score functions, such as the "Normal scores" that corresponds to $\varphi = \Phi^{-1}$, Φ being the d.f. of N(0, 1) r.v.. However, this is balanced by requiring (a), (c1) and (e). Note that (b) and (c1) together imply (2.3.5) with $d_{ni} = w_{ni}$, $1 \leq i \leq n$. Moreover, Theorem 3.2.2 does not require anything like (e).

Claim 3.2.1. (a) *implies that* f *is* $Lip(1/2)$.

First, from Hájek – Sĭdák (1967), pp 19–20, we recall that (a) implies that $f(x) \longrightarrow 0$ as $x \longrightarrow \pm \infty$. Now, absolute continuity and nonnegativity of f implies that

$$|f(x) - f(y)| \leq \int_x^y |(\dot{f}/f)| dF, \qquad x < y.$$

Therefore, by the Cauchy–Schwarz inequality, for $x < y$,

$$\text{(i)} \qquad |f(x) - f(y)| \leq \{\int_x^y (\dot{f}/f)^2 dF \cdot [F(y) - F(x)]\}^{1/2}$$

$$\text{(ii)} \qquad \leq I^{1/2}(f).$$

Letting $y \longrightarrow \infty$ in (ii) yields

$$\text{(iii)} \qquad \|f\|_\infty \leq I^{1/2}(f).$$

Now (i) and (iii) together imply

$$|f(x) - f(y)| \leq I^{1/2}(f) \{\int_x^y f(t)\, dt\}^{1/2} \leq I^{3/4}(f)\, (y-x)^{1/2}.$$

A similar inequality holds for $x > y$, thereby giving

$$|f(x) - f(y)| \leq I^{3/4}(f)\, |y-x|^{1/2}, \qquad \forall\ x, y \in \mathbb{R},$$

and proving the claim. Consequently, (a) *implies* (**F1**).

Note that f can be uniformly continuous, bounded, positive a.e., yet need not satisfy $I(f) < \infty$. For example, consider

$$\begin{aligned} f(x) &:= (1-x)/2,\ 0 \leq x \leq 1 \\ &:= (x-2j+1)/2^{j+2},\ \ 2j-1 \leq x \leq 2j \\ &:= (2j+1-x)/2^{j+2},\ \ 2j \leq x \leq 2j+1,\ \ j \geq 1; \\ f(x) &:= f(-x), \qquad x \leq 0. \end{aligned}$$

The above discussion shows that both Theorems 3.2.2 and 3.2.3 are needed. Neither displaces the other. If one is interested in the a.u.l. property of, say, Normal scores type rank statistics, then Theorem 3.2.3 gives an answer. On the other hand if one is interested in the a.u.l. property of,

say, the Wilcoxon type rank statistics, then Theorem 3.2.2 provides a better result.

The proof of Theorem 3.2.3 uses contiguity and projection technique *a la* Hájek (1962) to approximate $T_d(\varphi, u)$ for each fixed u. Then condition (e) implies the monotonicity of $T_d(\varphi, .)$ which yields the uniformity with respect to u. Such a proof is harder to extend to the case where u and c_{ni} are $p \times l$ vectors; this has been done by Jurečková (1971).

The proof of Theorem 3.2.2 exploits the monotonicity inherent in the w.e.p.'s Y_d and certain smoothness properties of F. It would be desirable to *extend this proof to include unbounded* φ. □

We now return to Theorem 3.2.1 with general $\{F_{ni}\}$. We wish to state an a.u.l. theorem for $\{Z_d\}$ and $\{T_d(\varphi, .)\}$ under general $\{F_{ni}\}$. Theorem 3.2.1 still does not quite do it because there is $\mathbf{u}$ in μ_w-expressions. We need to carry out an expansion of these terms in order to recover a term that is linear in $\mathbf{u}$. To that effect we have

Lemma 3.2.2. *In addition to the assumptions of Theorem* 3.2.1, *suppose that*

$$\text{(30)} \qquad n^{-1/2} \Sigma_i \|c_{ni}\| = 0(1).$$

Then, $\forall\ 0 < B < \infty$,

$$\text{(31)} \qquad \sup_{0 \le t \le 1, \|\mathbf{u}\| \le B} |n^{1/2}(HH_{n\mathbf{u}}^{-1}(t) - t) + Y_1(t, 0) + \mathbf{u}'\nu_1(t)| = o_p(1)$$

where Y_1, ν_1 etc. *are* Y_d, ν_d *of* (2.3.1), (2.3.8) *with* $d_{ni} \equiv n^{-1/2}$. *Consequently,*

$$\text{(32)} \qquad \sup_{0 \le t \le 1, \|\mathbf{u}\| \le B} |n^{1/2}(HH_{n\mathbf{u}}^{-1}(t) - t)| = O_p(1).$$

Proof. Write $Y_1(\cdot)$, $\mu_1(\cdot)$ for $Y_1(\cdot, 0)$, $\mu_1(\cdot, 0)$, respectively. Let I denote the identity function and set $\Delta_{n\mathbf{u}} := n^{1/2}(H_{n\mathbf{u}}H_{n\mathbf{u}}^{-1} - I)$. Then,

$$\begin{aligned}\text{(33)} \quad n^{1/2}(HH_{n\mathbf{u}}^{-1} - I) &= n^{1/2}(HH_{n\mathbf{u}}^{-1} - H_{\mathbf{u}}H_{n\mathbf{u}}^{-1} + H_{\mathbf{u}}H_{n\mathbf{u}}^{-1} - H_{n\mathbf{u}}H_{n\mathbf{u}}^{-1}) + \Delta_{n\mathbf{u}} \\ &= -[\mu_1(HH_{n\mathbf{u}}^{-1}, \mathbf{u}) - \mu_1(HH_{n\mathbf{u}}^{-1}) - \mathbf{u}'\nu_1(HH_{n\mathbf{u}}^{-1})] + \Delta_{n\mathbf{u}} \\ &\quad - \mathbf{u}'[\nu_1(HH_{n\mathbf{u}}^{-1}) - \nu_1] - \mathbf{u}'\nu_1 - Y_1 \\ &\quad - [Y_1(HH_{n\mathbf{u}}^{-1}, \mathbf{u}) - Y_1(HH_{n\mathbf{u}}^{-1})] - [Y_1(HH_{n\mathbf{u}}^{-1}) - Y_1].\end{aligned}$$

Now, note that $\sup_{t,u} |\Delta_{nu}(t)| \le n^{-1/2}$. Hence

$$(34)\quad \sup_{t,u}|n^{1/2}(HH_{nu}^{-1}(t) - t) + Y_1(t) + u'\nu_1(t)|$$

$$\le \sup_{t,u}|\mu_1(t, u) - \mu_1(t) - u'\nu_1(t)| + B \sup_t \|\nu_1(HH_{nu}^{-1}(t)) - \nu_1(t)\|$$

$$+ \sup_{t,u}| Y_1(t, u) - Y_1(t)| + \sup_{t,u}|W_1^*(HH_{nu}^{-1}(t)) - W_1^*(t)|,$$

where we have used the fact that $Y_1(t) = W_1^*(t)$ of (2.2a.33). The first term on the r.h.s. of (34) tends to zero by Lemma 2.3.1 when applied with $d_{ni} \equiv n^{-1/2}$. The third term tends to zero in probability by (2.3.25) applied with $d_{ni} \equiv n^{-1/2}$. To show that the other two terms go to zero in probability, use Lemma 3.2.1, (2.2a.35) and an analogue of (24) for ν_1 and an argument similar to the one that yielded (23) and (26) above. Thus we have (31). Since $\sup_{t,u} | Y_1(t, 0) + u'\nu_1(t)| = O_p(1)$, (32) follows. □

Lemma 3.2.3. *In addition to the assumptions of Theorem* 3.2.1 *and* (30), *suppose that for every* $0 < k < \infty$,

$$(35)\qquad \max_i \sup_{|t-s| \le kn^{-1/2}} n^{1/2}|L_{ni}(t) - L_{ni}(s) - (t-s)\ell_{ni}(s)| = o_p(1)$$

where $L_{ni} := F_{ni} H^{-1}$, $\ell_{ni} := f_{ni}(H^{-1})/h(H^{-1})$, $1 \le i \le n$ *with* $h := n^{-1}\sum_{i=1}^{n} f_{ni}$.

Moreover, suppose that, with $\tilde{w}(t) := n^{-1}\sum_{i=1}^{n} w_{ni}\, \ell_{ni}(t)$, $0 \le t \le 1$,

$$(36)\qquad \sup_{0\le t\le 1} n^{1/2}|\tilde{w}(t)| = O(1).$$

Then, $\forall\ 0 < B < \infty$,

$$(37)\qquad \sup|\mu_w(HH_{nu}^{-1}(t)) - \mu_w(t) + \{Y_1(t) + u'\nu_1(t)\}n^{1/2}\, \tilde{w}(t)| = o_p(1)$$

where $\mu_w(t)$, $Y_1(t)$ *stand for* $\mu_w(t, 0)$, $Y_1(t, 0)$, *respectively, and where the supremum is being taken over* $0 \le t \le 1$, $u \in \mathcal{N}(B)$.

Proof. Let $M_u := \mu_w(HH_{nu}^{-1}) - \mu_w$. From (32) it follows that $\forall\ \epsilon > 0$ $\exists\ K_\epsilon$ and $N_{1\epsilon}$ such that

$$(38)\qquad P(A_n^\epsilon) \ge 1 - \epsilon, \qquad n \ge N_{1\epsilon},$$

where

$$A_n^\epsilon = [\sup_{t,u} |HH_{nu}^{-1}(t) - t| \le K\epsilon n^{-1/2}].$$

By assumption (35), there exists $N_{2\epsilon}$ such that $n \ge N_{2\epsilon}$ implies

$$\max_i \sup_{|t-s| \le K\epsilon n^{-1/2}} n^{1/2} |L_i(t) - L_i(s) - (t-s)\ell_i(s)| < \epsilon. \tag{39}$$

Define

$$Z_{ui}^\epsilon := \{L_i(HH_{nu}^{-1}) - L_i - [HH_{nu}^{-1} - I]\, \ell_i\}\, I(A_n^\epsilon), \qquad 1 \le i \le n.$$

In view of (39) and (38),

$$P(\max_i \sup_{t,u} n^{1/2} |Z_{ui}^\epsilon(t)| > \epsilon) < \epsilon, \qquad n \ge N_{1\epsilon} \vee N_{2\epsilon} =: N_\epsilon. \tag{40}$$

Moreover,

$$\begin{aligned} M_u &= M_u\, I(A_n^\epsilon) + M_u\, I((A_n^\epsilon)^c) \\ &= \Sigma_i\, w_i\, Z_{ui}^\epsilon + Z_{u0}^\epsilon + n^{1/2}[HH_{nu}^{-1} - I]\, n^{1/2}\, \tilde{w}, \end{aligned} \tag{41}$$

where

$$Z_{u0}^\epsilon := \{M_u - n^{1/2}[HH_{nu}^{-1} - I] \cdot n^{1/2}\, \tilde{w}\}\, I((A_n^\epsilon)^c).$$

Note that

$$P(\sup_{t,u} |Z_{u0}^\epsilon| \ne 0) \le P((A_n^\epsilon)^c) < \epsilon, \qquad n > N_\epsilon. \tag{42}$$

By the C-S inequality, **(N3)** and (40),

$$P(\sup_{t,u} |\Sigma_i\, w_i\, Z_{ui}^\epsilon(t)| > \epsilon) < \epsilon, \quad n > N_\epsilon. \tag{43}$$

Hence, (37) follows from (43), (42), (41), Lemma 3.2.2 and (36). □

We combine Theorem 3.2.1, Lemmas 3.2.2 and 3.2.3 to obtain the following

Theorem 3.2.4. *Under the notation and assumptions of Theorem* 3.2.1, *Lemmas* 3.2.2 *and* 3.2.3, $\forall\ 0 < B < \infty$,

$$\sup |Z_d(t, u) - Z_d(t, 0) - u' \Sigma_i \left(d_{ni} - \tilde{d}_n(t)\right) c_{ni}\, q_{ni}(t)| = o_p(1), \tag{44}$$

(45) $\sup|T_d(\varphi, \mathbf{u})-T_d(\varphi, 0) + \mathbf{u}'\int \Sigma_i\,(d_{ni}-\tilde{d}_n(t))c_{ni}\,q_{ni}(t)\,d\varphi(t)| = o_p(1),$

where the supremum in (44) *is over* $0\le t\le 1$, $\|\mathbf{u}\|\le B$, *in* (45) *over* $\varphi\in\mathscr{C}$, $\|\mathbf{u}\|\le B$, *and where* $\tilde{d}_n(t) := n^{-1}\,\Sigma_i\, d_{ni}\,\ell_{ni}(t)$, $q_{ni} := f_{ni}(H^{-1}(t))$, $0\le t\le 1$, $1\le i\le n$.

Proof. Let $\rho(t) := \Sigma_i\,(d_i - \tilde{d}(t))c_iq_i(t)$. Note that the fact that $n^{-1}\Sigma_i\,\ell_i(t)\equiv 1$ implies that $\rho(t) = \Sigma_i\,(w_i - \tilde{w}(t))c_iq_i(t)$, where $\{w_i\}$ are as in (6). From (7), (8) and (9),

$$\text{(46)}\quad \text{l.h.s. }(44) = \sup_{0\le t\le 1,\,\|\mathbf{u}\|\le B} |Z_w(t, \mathbf{u}) - Z_w(t, 0) - \mathbf{u}'\rho(t)|$$

$$\le 4\max_i|w_i| + \sup_{0\le t\le 1,\,\|\mathbf{u}\|\le B} |\mathcal{T}_w(t, \mathbf{u}) - \mathcal{T}_w(t, 0) - \mathbf{u}'\rho(t)|.$$

Now, from Theorem 3.2.1 and Lemma 3.2.3, uniformly in $0\le t\le 1$, $\|\mathbf{u}\|\le B$,

$$\text{(47)}\quad \sup|T^*_w(t, \mathbf{u}) - Y_w(t) + \{Y_1(t) + \nu_1(t)\,\mathbf{u}\}\,n^{1/2}\,\tilde{w}(t)| = o_p(1),$$

where $Y_d(t)$ stands for $Y_d(t, 0)$ for arbitrary weights $\{d_{ni}\}$. Therefore,

$$\sup|\mathcal{T}_w(t, \mathbf{u}) - \mathcal{T}_w(t, 0) - \mathbf{u}'\rho(t)|$$

$$= \sup|\tilde{T}_w(t, \mathbf{u}) - \tilde{T}_w(t, 0) + \mu_w(t, \mathbf{u}) - \mu_w(t, 0) - \mathbf{u}'\rho(t)|$$

$$\le \sup|\tilde{T}_w(t, \mathbf{u}) - \tilde{T}_w(t, 0) + \mathbf{u}'\nu_1(t)\,n^{1/2}\,\tilde{w}(t)|$$

$$+ \sup|\mu_w(t, \mathbf{u}) - \mu_w(t, 0) - \mathbf{u}'\nu_w(t)| = o_p(1),$$

by (47) and Lemma 2.3.1 and the fact that $\rho(t) = \nu_w(t)-\nu_1(t)\,n^{1/2}\,\tilde{w}(t)$. This completes the proof (44). The proof of (45) follows from (44) in the same fashion as does that of (27) from (26). □

Remark 3.2.3. As in Remark 2.2a.3, suppose we strengthen **(N3)** to require

$$\text{(B1)}\qquad n\max_i w_{ni}^2 = O(1), \qquad \tau_w^2 = 1.$$

Then **(C*)** and (36) are *a priori* satisfied by L_w. □

Remark 3.2.4. If one is interested in the i.i.d. case only, then Theorem 3.2.2 gives a better result than Theorem 3.2.4. □

3.3. A.U.L. OF LINEAR SIGNED RANK STATISTICS

In this section our aim is to prove analogs of Theorems 3.2.2 and 3.2.4 for the signed rank processes $\{\overset{+}{T}_d(\varphi, \mathbf{u}), \mathbf{u}\in\mathbb{R}^p\}$, using as many results from the previous sections as possible. Many details are quite similar. Define, for $\mathbf{u}\in\mathbb{R}^p$, $0 \le t \le 1$, $x \ge 0$,

$$
\begin{aligned}
(1)\qquad Z_d^+(t, \mathbf{u}) &:= \Sigma_i\, d_{ni}\, I(R_{i\mathbf{u}}^+ \le nt)\, s(X_{ni} - \mathbf{c}_{ni}'\mathbf{u}),\\
J_{n\mathbf{u}}(x) &:= n^{-1}\Sigma_i\, I(|X_{ni} - \mathbf{c}_{ni}'\mathbf{u}| \le x) = H_{n\mathbf{u}}(x) - H_{n\mathbf{u}}(-x),\\
J_{\mathbf{u}}(x) &:= n^{-1}\, \Sigma_i\, [F_{ni}(x+\mathbf{c}_{ni}'\mathbf{u}) - F_{ni}(-x+\mathbf{c}_{ni}'\mathbf{u})] = H_{\mathbf{u}}(x) - H_{\mathbf{u}}(-x),\\
\mathcal{T}_d^+(t, \mathbf{u}) &:= \Sigma_i\, d_{ni} I(|X_{ni} - \mathbf{c}_{ni}'\mathbf{u}| \le J_{n\mathbf{u}}^{-1}(t))\, s(X_{ni} - \mathbf{c}_{ni}'\mathbf{u}),\\
S_d^+(t, \mathbf{u}) &:= \Sigma\, d_{ni}\, I(|X_{ni} - \mathbf{c}_{ni}'\mathbf{u}| \le J^{-1}(t))\, s(X_{ni} - \mathbf{c}_{ni}'\mathbf{u}),\\
\mu_d^+(t, \mathbf{u}) &:= \Sigma_i\, d_{ni}\, \mu_{ni}^+(t, \mathbf{u}) = E\, S_d^+(t, \mathbf{u}),\\
\mu_{ni}^+(t, \mathbf{u}) &:= F_{ni}(J^{-1}(t)+\mathbf{c}_{ni}'\mathbf{u}) + F_{ni}(-J^{-1}(t)+\mathbf{c}_{ni}'\mathbf{u}) - 2F_{ni}(\mathbf{c}_{ni}'\mathbf{u}), \quad 1\le i\le n.
\end{aligned}
$$

In the above and sequel, J and J_n stand for J_0 and J_{n0}, respectively. We also need,

$$(2)\qquad \mathcal{Y}_d^+(t, \mathbf{u}) := S_d^+(t, \mathbf{u}) - \mu_d^+(t, \mathbf{u}),$$

and

$$(3)\qquad \tilde{T}_d^+(t, \mathbf{u}) := \mathcal{T}_d^+(t, \mathbf{u}) - \mu_d^+(t, \mathbf{u}), \qquad 0 \le t \le 1,\ \mathbf{u} \in \mathbb{R}^p.$$

Analogous to (3.2.11), we have the basic decomposition: For $0\le t\le 1$, $\mathbf{u}\in\mathbb{R}^p$,

$$(4)\qquad \tilde{T}_d^+(t, \mathbf{u}) = \mathcal{Y}_d^+(JJ_{n\mathbf{u}}^{-1}(t), \mathbf{u}) + \mu_d^+(JJ_{n\mathbf{u}}^{-1}(t), \mathbf{u}) - \mu_d^+(t, \mathbf{u}),$$

Now, note that, w.p. 1, for all $0 \le t \le 1$, $\mathbf{u} \in \mathbb{R}^p$,

$$(5)\qquad \mathcal{Y}_d^+(t, \mathbf{u}) = \mathcal{Y}_d(HJ^{-1}(t), \mathbf{u}) + \mathcal{Y}_d(H(-J^{-1}(t)), \mathbf{u}) - 2\, \mathcal{Y}_d(H(0), \mathbf{u}),$$

where $\mathcal{Y}_d$ is as in (2.3.1). Therefore, by Theorem 2.3.1 (see (2.3.25)), under the assumptions of that theorem and strictly increasing nature of J and H,

$$(6) \qquad \sup_{t,u} |\, Y_d^+(t, u) - Y_d^+(t, 0)| = o_p(1).$$

One also has, in view of the continuity of $\{F_{ni}\}$, a relation like (5) between μ_d^+ and μ_d. Thus by Lemma 2.3.1, under the assumptions there,

$$(7) \qquad \sup_{t,u} |\mu_d^+(t, u) - \mu_d^+(t, 0) - u' \nu_d^+(t)| = o(1)$$

where

$$(8) \qquad \nu_d^+(t) := \Sigma\, d_{ni}\, c_{ni}\, [f_{ni}(J^{-1}(t)) + f_{ni}(-J^{-1}(t)) - 2f_{ni}(0)], \quad 0 \le t \le 1.$$

We also have an anlogue of Lemma 3.2.1:

Lemma 3.3.1. *Without any assumption except* (2.2a.34),

$$(9) \qquad \sup_{0 \le x \le \infty} |J_n(x) - J(x)| \longrightarrow 0 \text{ a.s.}$$

If, in addition, (2.3.4) *and* (3.2.15) *hold, then*

$$(10) \qquad \sup_{0 \le x \le \infty,\, \|u\| \le B} |J_{nu}(x) - J_u(x)| \longrightarrow 0 \text{ a.s..}$$

Using this lemma, arguments like those in Theorem 3.2.1 and the above discussion, one obtains

Theorem 3.3.1. *Suppose that* $\{X_{ni}, F_{ni}\}$ *satisfy* (2.2a.34), (2.3.3b) *and that* $\{d_{ni}, c_{ni}\}$ *satisfy* (N1), (N2), (2.3.4) *and* (2.3.5). *In addition, assume that*

$$(11) \qquad \lim_{\delta \to 0} \limsup_n \max_i \sup_{|J(x) - J(y)| < \delta} |f_{ni}(x) - f_{ni}(y)| = 0$$

and that H *is strictly increasing for every* n. *Then, for every* $0 < B < \infty$,

$$(12) \qquad \sup_{0 \le t \le 1,\, \|u\| \le B} |\tilde{T}_d^+(t, u) - Y_d^+(t, 0) - \mu_d^+(JJ_{nu}^{-1}(t), 0) + \mu_d^+(t, 0)| = o_p(1). \qquad \square$$

We remark here that (11) implies (3.2.12).

Next, note that if $\{F_i\}$ are symmetric about 0, then

$$(13) \qquad \mu_d^+(t, 0) = 0, \qquad 0 \le t \le 1, \qquad n \ge 1.$$

Upon combining (13), (12) with (7) one obtains

Theorem 3.3.2. *In addition to the assumptions of Theorem* 3.2.1, *suppose that* $\{F_{ni}, 1 \le i \le n\}$ *are symmetric about* 0.
Then, for every $0 < B < \infty$,

$$(14)\qquad \sup_{0\le t\le 1,\, \|u\|\le B} |Z_d^+(t, u) - Z_d^+(t, 0) - u'\Sigma_i\, d_{ni}c_{ni}\, \nu_{ni}^+(t)| = o_p(1),$$

$$(15)\qquad \sup_{\varphi\in \mathscr{C},\|u\|\le B} |T_d^+(\varphi, u) - T_d^+(\varphi, 0) + u'\Sigma_i\, d_{ni}c_{ni} \int_0^1 \nu_{ni}^+(t)\, d\varphi^+(t)| = o_p(1),$$

where

$$\nu_{ni}^+(t) := 2[f_{ni}(J^{-1}(t) - f_{ni}(0)], \quad 1 \le i \le n, \quad 0 \le t \le 1.$$

Proof. Using a relation like (3.2.5) between R_{iu}^+ and J_{nu}, one obtains, as in (3.2.9),

$$(16)\qquad \sup_{t,u} |Z_d^+(t, u) - \mathscr{V}_d^+(t, u)| \le 2\max_i |d_i| = o(1), \qquad \text{by } \mathbf{(N2)}.$$

Thus (13) follows from (16), (12), (11) and (7). Conclusion (15) follows from (13) in the same way as (3.2.27) follows from (3.2.26). □

Because of the importance of the i.i.d. symmetric case, we specialize the above theorem to yield

Corollory 3.3.1. *Let* F *be a d.f., symmetric around zero, satisfying* **(F1)**, **(F2)** *and let* $X_{n1}, \ldots, X_{nn}$ *be i.i.d.* F. *In addition, assume that* $\{d_{ni}, c_{ni}\}$ *satisfy* **(N1)**, **(N2)**, (2.3.4) *and* (2.3.5). *Then, for every* $0 < B < \infty$,

$$(17)\qquad \sup_{0\le t\le 1,\, \|u\|\le B} |Z_d^+(t, u) - Z_d^+(t, 0) - u'\Sigma_i\, d_{ni}\, c_{ni}\, q^+(t)| = o_p(1),$$

$$(18)\qquad \sup_{\varphi\in \mathscr{C}, u\in \mathscr{N}(B)} |T_d^+(\varphi,u) - T_d^+(\varphi,0) + \Sigma_i\, d_{ni}c_{ni}'\, u\int_0^1 q^+(t)d\varphi^+(t)| = o_p(1),$$

where $q^+(t) := 2[f(F^{-1}((t+1)/2)) - f(0)],\ 0 \le t \le 1.$ □

Remark 3.3.1. Van Eeden (1972) proved an analogue of (18) without the supremum over φ, but for square integrable φ's. She also needs conditions like those in Theorem 3.2.3 above. Thus Remark 3.2.1 is equally applicable here when comparing Corollory 3.2.1 with Van Eeden's results. □

Now, we return to Theorem 3.3.1 and expand the μ_d^+-terms further so as to recover an extra linearity term. Define, for $0 \le t \le 1$, $u \in \mathbb{R}^p$,

$$(19) \qquad \overset{*}{Y}_d(t, \mathbf{u}) := \Sigma_i\, d_{ni}[I(|X_{ni} - \mathbf{c}_{ni}\,\mathbf{u}| \le J^{-1}(t)) - F^{+}_{i\mathbf{u}}(J^{-1}(t))]$$

$$\overset{*}{\nu}_d(t) := \Sigma_i\, d_{ni}\, \mathbf{c}_{ni}\, [f_{ni}(J^{-1}(t)) - f_{ni}(-J^{-1}(t))]$$

where

$$F^{+}_{i\mathbf{u}}(x) := F_{ni}(x + \mathbf{c}_i'\mathbf{u}) - F_{ni}(-x + \mathbf{c}_i'\mathbf{u}), \qquad x \ge 0.$$

Note the relation: For arbitrary $\{d_{ni}\}$,

$$(20) \qquad \overset{*}{Y}_d(t, \mathbf{u}) \equiv Y_d(HJ^{-1}(t), \mathbf{u}) - Y_d(H(-J^{-1}(t)), \mathbf{u}).$$

From (20) and (2.3.25) applied with $d_{ni} = n^{-1/2}$, we obtain

$$(21) \qquad \sup_{t,\mathbf{u}} |\overset{*}{Y}_1(t, \mathbf{u}) - \overset{*}{Y}_1(t, 0)| = o_p(1).$$

Note that in the case $d_{ni} \equiv n^{-1/2}$, (2.3.5) reduces to (3.2.30).
Next, under (11) and (2.3.5), just as (3.2.24),

$$(22) \qquad \lim_{\delta\to 0} \limsup_n \sup_{|t-s|<\delta} \|\overset{*}{\nu}_d(t) - \overset{*}{\nu}_d(s)\| = 0,$$

for the given $\{d_{ni}\}$ and for $d_{ni} \equiv n^{-1/2}$.
Using (21), (22) and calculations similar to those done in the proof of Lemma 3.2.2, we obtain

Lemma 3.3.2. *Under the conditions of Theorem* 3.2.1 *and* (3.2.30)

$$(23) \qquad \sup_{t,\mathbf{u}} |n^{1/2}(JJ^{-1}_{n\mathbf{u}}(t) - t) + \overset{*}{Y}_1(t, 0) + \mathbf{u}'\,\overset{*}{\nu}_1(t)\,| = o_p(1).$$

Consequently,

$$(24) \qquad \sup_{t,\mathbf{u}} |n^{1/2}(JJ^{-1}_{n\mathbf{u}}(t) - t)| = O_p(1). \qquad \square$$

Similarly arguing as in Lemma 3.2.3, we obtain the following Lemma 3.3.3. In it $\mu^{+}_d(t)$, $\mu^{+}_1(t)$ etc. stand for $\mu^{+}_d(t, 0)$, $\mu^{+}_1(t, 0)$ etc. of (1).

***Lemma* 3.3.3.** *In addition to the assumptions of Theorem* 3.2.1, (3.2.30) *assume that for every* $0 < k < \infty$,

$$(25) \qquad \max_i \sup_{|t-s|\le kn^{-1/2}} n^{1/2}\, |\mu^{+}_{ni}(t) - \mu^{+}_{ni}(s) - (t-s)\ell^{+}_{ni}(s)| = o(1)$$

where $\{\mu^+_{ni}\}$ *are as in* (1),

$$(26)\qquad \ell^+_{ni}(s) := [f_{ni}(J^{-1}(s)) - f_{ni}(-J^{-1}(s))] \,/\, h^+(J^{-1}(s)), \qquad 0 \le s \le 1,$$

$$h^+(x) := n^{-1}\,\Sigma_i\,[f_{ni}(x) - f_{ni}(-x)], \qquad x \ge 0.$$

Moreover, with $\tilde{d}^+_n(t) := n^{-1}\,\Sigma_i\, d_{ni}\ell^+_{ni}(t),\ 0 \le t \le 1$, *assume that*

$$(27)\qquad \sup_{0\le t\le 1} |n^{1/2}\,\tilde{d}^+_n(t)| = O(1).$$

Then,

$$(28)\qquad \sup|\mu^+_d(JJ^{-1}_{nu}(t)) - \mu^+_d(t) + \{\overset{*}{Y}_1(t) + u'\overset{*}{\nu}_1(t)\}n^{1/2}\,\tilde{d}^+_n(t)| = o_p(1),$$

where the supremum is taken over the set $0\le t\le 1,\ \|u\|\le B$. □

Finally, an analogue of Theorem 3.2.3 is

Theorem 3.3.3. *Under the assumptions of Theorem* 3.3.1, (3.2.30), (25) *and* (27), *for every* $0 < B < \infty$,

$$(29)\qquad \sup_{0\le t\le 1,\|u\|\le B} |Z^+_d(t, u) - Z^+_d(t, 0) - u'[\nu^+_d(t) - \overset{*}{\nu}_1(t)n^{1/2}\,\tilde{d}^+_n(t)]| = o_p(1),$$

$$(30)\quad \sup\ |T^+_d(\varphi,u) - T^+_d(\varphi,0) + u'\int_0^1[\nu^+_d(t) - \overset{*}{\nu}_1(t)n^{1/2}\tilde{d}^+_n(t)]\,d\varphi^+(t)| = o_p(1),$$

where the supremum in (30) *is over* $\varphi \in \mathscr{C}$, $\|u\|\le B$. □

Remark 3.3.2. Unlike the case in Theorem 3.2.3, there does not appear to be a nice simplification of the term $\nu^+_d - \overset{*}{\nu}_1\, n^{1/2}\,\tilde{d}^+_n$. However, it can be rewritten as follows:

$$\nu^+_d(t) - \overset{*}{\nu}_1(t)n^{1/2}\,\tilde{d}^+_n(t) = \Sigma_i\, d_i c_i\,[f_i(J^{-1}(t)) + f_i(-J^{-1}(t)) - 2f_i(0)]$$
$$+ \Sigma_i\,(d_i - \tilde{d}^+_n(t))\, c_i\,[f_i(J^{-1}(t)) - f_i(-J^{-1}(t))].$$

This representation is somewhat revealing in the following sense. The first term is due to the shift $u'c_i$ in the r.v. X_i and the second term is due to the nonidentical and asymmetric nature of the distribution of X_i, $1 \le i \le n$. □

Remark 3.3.3. If one is interested in the symmetric case or in the i.i.d. symmetric case then Theorem 3.3.2 and Corollary 3.3.1, respectively, give better results than Theorem 3.3.3. □

3.4. WEAK CONVERGENCE OF RANK AND SIGNED RANK W.E.P.'S.

Throughout this section we shall use the notation of Sections 3.2 – 3.3 with $u = 0$. Thus, e.g., $Z_d(t)$, $Z_d^+(t)$, etc. will represent $Z_d(t, 0)$, $Z_d^+(t, 0)$, etc. of (3.2.2) and (3.3.1), i.e., for $0 \le t \le 1$,

$$(1)\quad Z_d(t) = \Sigma_i\, d_{ni}\, I(R_{ni} \le nt), \qquad Z_d^+(t) = \Sigma_i\, d_{ni}\, I(R_{ni}^+ \le nt)s(X_{ni}),$$

$$\mathscr{V}_d(t) = \Sigma_i\, d_{ni}\, I(X_{ni} \le H_n^{-1}(t)), \qquad \mu_d(t) = \Sigma_i\, d_{ni}\, L_{ni}(t),$$

where R_{ni} (R_{ni}^+) is the rank of X_{ni} $(|X_{ni}|)$ among $X_{n1}, ..., X_{nn}$ $(|X_{n1}|, ..., |X_{nn}|)$.

We shall first prove the asymptotic normality of Z_d and Z_d^+ for a fixed t, say $t = v$, $0<v<1$. To begin with consider $Z_d(v)$. In the following theorem v is a fixed number in $(0, 1)$.

Theorem 3.4.1. *Suppose that* $\{X_{ni}\}$, $\{F_{ni}\}$, $\{L_{ni}\}$, L_d *are as in* (2.2a.33) and (2.2a.34). *Assume that* $\{d_{ni}\}$ *satisfy* (N1), (N2) *and that* H *is strictly increasing for each* n. *Also assume that*

$$(2)\qquad \lim_{\delta\to 0} \limsup_n\, [L_d(v + \delta) - L_d(v - \delta)] = 0,$$

and that there are nonnegative numbers $\ell_{ni}(v)$, $1 \le i \le n$, *such that for every* $0 < k < \infty$,

$$(3)\qquad \max_i \sup_{|t-s| \le kn^{-1/2}} n^{1/2}|L_{ni}(t) - L_{ni}(v) - (t-v)\ell_{ni}(v)| = o(1).$$

Denoting

$$(4)\quad \tilde{d}_n(v) := n^{-1}\, \Sigma_i\, d_{ni}\ell_{ni}(v), \quad \sigma_d^2(v) := \Sigma_i\, (d_{ni} - \tilde{d}_n(v))^2 L_{ni}(v)(1-L_{ni}(v)),$$

assume that

$$(5)\qquad n^{1/2}|\tilde{d}_n(v)| = O(1).$$

$$(6)\qquad \liminf_n \sigma_d^2(v) > 0.$$

Then,

$$\{\sigma_d(v)\}^{-1}\{Z_d(v) - \mu_d(v)\} \xrightarrow[d]{} N(0, 1).$$

The proof of Theorem 3.4.1 is a consequence of the following *three* lemmas. In these lemmas the setup is the same as in Theorem 3.4.1.

Lemma 3.4.1. *Under the sole assumption of* (2.2a.34),

$$\sup_{0\le t\le 1} |HH_n^{-1}(t) - t| = o_p(1).$$

Proof. Upon taking $u = 0$ in (3.2.19), one obtains

$$\sup_{0\le t\le 1} |HH_n^{-1}(t) - t| \le \sup_{-\infty\le x\le +\infty} |H_n(x) - H(x)| + n^{-1} = o_p(1),$$

by (3.2.14) of Lemma 3.2.1. □

Lemma 3.4.2. *Let* $Y_d(t)$ denote the $Y_d(t, 0)$ of (2.3.1). *Then, under* (3), *for every* $\epsilon > 0$,

$$\lim_{\delta\to 0} \limsup_n P\Big(\sup_{|t-v|<\delta} |Y_d(t) - Y_d(v)| > \epsilon\Big) = 0$$

Proof. Apply Lemma 2.2a.2 to $\eta_{ni} = H(X_{ni})$, $G_{ni} = L_{ni}$, to obtain that $Y_d \equiv W_d$ of that lemma and that

$$P\Big(\sup_{|t-v|<\delta} |Y_d(t) - Y_d(v)| > \epsilon\Big)$$

$$\le \kappa\epsilon^{-2}[L_d(v + \delta) - L_d(v - \delta)]^2 + P(|Y_d(v - \delta) - Y_d(v)| > \epsilon/2)$$

$$+ P(|Y_d(v + \delta) - Y_d(v - \delta)| > \epsilon/4)$$

$$\le (\kappa + 20)\epsilon^{-2} [L_d(v + \delta) - L_d(v - \delta)], \qquad \text{(by Chebyshev).}$$

The Lemma now follows from the assumption (3). □

Lemma 3.4.3. *Under* (3), *for every* $\epsilon > 0$,

$$\limsup_n P(|Y_d(HH_n^{-1}(v)) - Y_d(v)| > \epsilon) = 0.$$

Proof. Follows from Lemmas 3.4.1 and 3.4.2 . □

Remark 3.4.1. Lemmas 3.4.2 could be deduced from Corollary 3.3.1 which gives the tightness of the process Y_d under stronger condition (C*). But here we are interested in the behavior of Y_d only in the neighborhood of one point v and the above lemma proves the contnuity of Y_d at the point v at which (3) holds. Similarly, many of the approximations that follow could of course be deduced from proofs of Theorems 3.2.1 and 3.2.2. But these theorems obtain results uniformly in $0\le t\le 1$ under rather stronger conditions than would be needed in the present case. Of course various decompositions used in their proofs will be useful here also. □

Proof of Theorem 3.4.1. In view of (3.2.9) and (**N2**), it suffices to prove that $\{\sigma_d(v)\}^{-1}\,\tilde{T}_d(v) \underset{d}{\longrightarrow} N(0,1)$, where

(7) $$\tilde{T}_d(v) = \{\, \mathcal{V}_d(v) - \mu_d(v)\}.$$

But, from (3.2.11) applied with $\mathbf{u} = \mathbf{0}$,

$$\tilde{T}_d(v) = Y_d(HH_n^{-1}(v)) + \mu_d(HH_n^{-1}(v)) - \mu_d(v), \quad \text{w.p. } 1.$$

(8) $$= Y_d(v) + o_p(1) + \mu_d(HH_n^{-1}(v)) - \mu_d(v), \qquad \text{by (6)}.$$

Apply the identity (3.2.33) with $\mathbf{u} = \mathbf{0}$ and Lemma 3.4.3 with $d_i \equiv n^{-1/2}$ to obtain,

(9) $$n^{1/2}[HH_n^{-1}(v) - v] = -Y_1(HH_n^{-1}(v)) + o_p(1) = -Y_1(v) + o_p(1).$$

Since $Y_1(v) \underset{d}{\longrightarrow} N(0,\, v(1-v))$, $|Y_1(v)| = O_p(1)$. Again, argue as for (3.2.37) with $\mathbf{u} \equiv \mathbf{0}$, $t \equiv v$ (i.e., without the supremum on the l.h.s. and with $\mathbf{u} \equiv \mathbf{0}$, $t \equiv v$), to conclude that

(10) $$\mu_d(HH_n^{-1}(v)) - \mu_d(v) = -Y_1(v)\, n^{1/2}\, \tilde{d}(v) + o_p(1).$$

Combine (9), (10) to obtain

(11) $$\tilde{T}_d(v) = Y_d(v) - n^{1/2}\,\tilde{d}(v)\; Y_1(v) + o_p(1)$$

$$= \sum_{i=1}^{n} (d_{ni} - \tilde{d}(v))\; \{I(X_{ni} \le H^{-1}(v)) - L_{ni}(v)\} + o_p(1).$$

The theorem now follows from (6) and the fact that $\{\sigma_d(v)\}^{-1}\cdot\{$leading term in the r.h.s. of (11)$\} \underset{d}{\longrightarrow} N(0,1)$ by the L-F CLT, in view of (**N1**) and (**N2**).□

Remark 3.4.2. If $\{F_{ni}\}$ have densities $\{f_{ni}\}$ then $\ell_{ni}(v)$ can be taken to be $f_{ni}(H^{-1}(v))/h(H^{-1}(v))$, just as in (3.2.34). However, if one is interested in the asymptotic normality of linear rank statistic corresponding to the jump score function, with jump at v, then we need $\{L_{ni}\}$ to be smooth only at that jump point.

The above Theorem 3.4.1 bears strong resemblance to Theorem 1 of Dupač–Hájek (1969). The assumptions (**N1**), (**N2**) and (4) correspond to (2.2a), (2.13) and (2.2a2) of Dupač-Hájek. Condition (3) above is not quite comparable to condition (2.12) Dupač-Hájek but it appears to be less restrictive. In any case, (2.12) and (2.13) together imply the boundedness of

$\{\ell_i(v)\}$ and hence the condition (5) above. Taken together, then, the assumptions of the above theorem are somewhat weaker than those of Dupač-Hájek. On the other hand, the conclusions of the Dupač-Hájek Theorem 1 are stronger than those of the above theorem in that it asserts not only $\{Z_d(v) - \mu_d(v)\}\sigma_d^{-1}(v) \Rightarrow N(0,1)$ but also that $E[\sigma_d^{-1}(v)(Z_d(v) - \mu_d(v))]^r \longrightarrow 0$, for $r = 1, 2$, as $n \longrightarrow \infty$. However, if one is only interested in the asymptotic normality of $\{Z_d(v)\}$ then the above theorem appears to be more desirable. Moreover, in view of the decomposition (3.2.11), the proof presented below makes the role played by conditions (3) and (4) clearer.

The assumption about H being strictly increasing is not really an assumption because, without loss of generality, one may assume that $\{F_i\}$ are not flat on a common interval. For, if all $\{F_i\}$ were flat on a common interval, then deletion of this interval would not change the distribution of $R_1, \ldots, R_n$ and hence of $\{Z_d\}$. □

Next, we turn to the asymptotic normality of $Z_d^+(v)$. Again, put $\mathbf{u} = \mathbf{0}$ in the definition (3.3.1) to obtain,

(12) $$\mathscr{V}_d^+(t) = \Sigma_i\, d_{ni}\, I(|X_{ni}| \le J_n^{-1}(t))s(X_{ni}),$$

$$\mu_{ni}^+(t) = F_{ni}(J^{-1}(t)) + F_{ni}(-J^{-1}(t)) - 2F_{ni}(0), \quad 1 \le i \le n$$

$$S_d^+(t) = \Sigma_i\, d_{ni} I(|X_{ni}| \le J^{-1}(t))s(X_{ni}), \quad 0 \le t \le 1,$$

$$\mu_d^+(t) = \Sigma_i\, d_{ni}\, \mu_{ni}^+(t), \quad 0 \le t \le 1. \qquad Y_d^+ = S_d^+ - \mu_d^+.$$

Like (3.2.9), we have

(13) $$\sup_{0 \le t \le 1} |Z_d^+(t) - \mathscr{V}_d^+(t)| \le 2 \max_i |d_i|.$$

Because of (N2), it suffices to consider $\mathscr{V}_d^+$ only. Observe that

$$Y_d^+(t) = Y_d(HJ^{-1}(t)) + Y_d(H(-J^{-1}(t))) - 2Y_d(H(0)),$$

where Y_d is as in (2.3.1). Rewrite

(14) $$Y_d^+(t) = \{Y_d(HJ^{-1}(t)) - Y_d(H(0))\} - \{Y_d(H(0)) - Y_d(-J^{-1}(t))\}$$

$$= Y_{d1}^*(t) - Y_{d2}^*(t), \quad \text{say.}$$

This representation motivates the following notation as it is required in the subsequent lemma. Let $p_i := F_i(0)$, $q_i := 1 - p_i$ and define for $0 \le t \le 1$,

$$
\begin{aligned}
(15) \qquad L_{i1}^{+}(t) &:= \{F_i(J^{-1}(t)) - p_i\}/q_i, & q_i > 0,\\
&= 0, & q_i = 0;\\
L_{i2}^{+}(t) &:= \{p_i - F_i(-J^{-1}(t))\}/p_i, & p_i > 0,\\
&= 0, & p_i = 0;\;\; 1 \le i \le n.
\end{aligned}
$$

Observe that $\mu_i^{+}(v) = q_i L_{i1}^{+}(v) - p_i L_{i2}^{+}(v)$, $1 \le i \le n$. Also define

$$
\begin{aligned}
(16) \qquad L_i^{+}(t) &:= q_i L_{i1}^{+}(t) + p_i L_{i2}^{+}(t) = P(|X_i| \le J^{-1}(t)), & 1 \le i \le n,\\
L_{d1}^{+}(t) &:= \Sigma_i\, d_i^2\, q_i\, L_{i1}^{+}(t), \;\; L_{i1}^{+}(t) := \Sigma_i\, d_i^2\, p_i\, L_{i2}^{+}(t), & 0 \le t \le 1.
\end{aligned}
$$

Argue as for the proof of Lemma 2.2a.2 and use the triangle and the Chebychev inequalitites to conclude

Lemma 3.4.4. *For every* $\epsilon > 0$ *and* $0 < v < 1$ *fixed,*

$$
\begin{aligned}
&P\Big(\sup_{|t-v|<\delta} |Y_{dj}^{*}(t) - Y_{dj}^{*}(v)| > \epsilon\Big)\\
(17) \qquad &\le (\kappa + 20)\epsilon^{-2}\,[L_{dj}^{+}(v + \delta) - L_{dj}^{+}(v - \delta)], \qquad j = 1, 2
\end{aligned}
$$

where κ *does not depend on* ϵ, δ *or any other underlying quantities.* □

Theorem 3.4.2. *Let* $X_{n1}, \ldots, X_{nn}$ *be independent r.v.'s with respective continuous d.f.'s* $F_{n1}, \ldots, F_{nn}$ *and* $d_{n1}, \ldots, d_{nn}$ *be real numbers. Assume that* $\{d_{ni}\}$ *satisfy* (N1), (N2). *In addition, assume the following.*

With $\{L_{dj}^{+}\}$ *as in* (16), *for* v *fixed in* (0, 1),

$$(18) \qquad \lim_{\delta \to 0} \limsup_n |L_{dj}^{+}(v + \delta) - L_{dj}^{+}(v - \delta)| = 0, \qquad j = 1, 2.$$

(19) *There exist numbers* $\{\ell_{ij}^{+}(v),\ 1 \le i \le n;\ j = 1, 2\}$ *such that for all* $0 < k < \infty$, $j = 1, 2$,

$$\max_i \sup_{|t-v| \le kn^{-1/2}} n^{1/2} |L_{ij}^{+}(t) - L_{ij}^{+}(v) - (t - v)\ell_{ij}^{+}(v)| = o(1).$$

With

$$(20) \qquad \tilde{d}_n^{+}(v) := n^{-1}\Sigma_i\, d_{ni}\{q_i \ell_{i1}^{+}(v) - p_i \ell_{i2}^{+}(v)\},$$

$$\tau^2(v) := \Sigma_i \{d_{ni}^2[L_{ni}^+(v)-\{\mu_{ni}^+(v)\}^2] + (\tilde{d}_n^+(v))^2 L_{ni}^+(v)(1-L_{ni}^+(v)) - 2d_{ni}\tilde{d}_n^+(v)\mu_{ni}^+(v)(1-L_{ni}^+(v))\},$$

(21a) $$\liminf_n \tau^2(v) > 0.$$

(21b) $$\limsup_n n^{1/2}|\tilde{d}_n^+(v)| < \infty.$$

Then,

(22) $$\{\tau(v)\}^{-1}[Z_d^+(v) - \mu_d^+(v)] \underset{d}{\longrightarrow} N(0, 1)$$

where μ_d^+ *is as in* (12).

Proof. The proof of this theorem is similar to that of Theorem 3.4.1 so we shall be brief. To begin with, by (13) and **(N2)** it suffices to prove that $\{\tau(v)\}^{-1}\tilde{T}_d^+(v) \underset{d}{\longrightarrow} N(0, 1)$, where $\tilde{T}_d^+(v) := \mathcal{V}_d^+(v) - \mu_d^+(v)$.

Apply Lemma 3.4.1 above to the r.v.'s $|X_{n1}|, \ldots, |X_{nn}|$, to conclude that

$$\sup_{0\le t\le 1} |J(J_n^{-1}(t)) - t| = o_p(1).$$

From this, (14), (17) and (18),

$$\tilde{T}_d^+(v) = \mathcal{Y}_d^+(JJ_n^{-1}(v)) + \mu_d^+(JJ_n^{-1}(v)) - \mu_d^+(v).$$

$$= \mathcal{Y}_d^+(v) + [\mu_d^+(JJ_n^{-1}(v)) - \mu_d^+(v)] + o_p(1).$$

Again, apply arguments like those that yielded (9) to $\{|X_{ni}|\}$ to obtain

$$n^{1/2}[JJ_n^{-1}(v) - v] = -\mathcal{Y}_1^*(v) + o_p(1),$$

where $\mathcal{Y}_1^*(v)$ is as in (3.3.19) with $t = v$ and $\mathbf{u} = \mathbf{0}$. Consequently,

$$\tilde{T}_d^+(v) = \mathcal{Y}_d^+(v) - n^{1/2}\tilde{d}_n^+(v)Y_1^*(v) + o_p(1) = K_d^+(v) + o_p(1)$$

where

$$K_d^+(v) = \mathcal{Y}_d^+(v) - n^{1/2}\tilde{d}_n^+(v)\mathcal{Y}_1^*(v)$$
$$= \Sigma_i \{d_{ni}[I(J(|X_{ni}|) \le v)\, s(X_{ni}) - \mu_{ni}^+(v)] - \tilde{d}_n^+(v)[I(J(|X_{ni}|) \le v) - L_{ni}^+(v)]\}.$$

Note that $\mathrm{Var}(K_d^+(v)) = \tau^2(v)$. The proof of the theorem is now completed by using the L-F CLT which is justified, in view of (N1), (N2), and (21a). □

Remark 3.4.3. Observe that if $\{F_i\}$ are symmetric about 0 then $\mu_i^+ \equiv 0 \equiv \tilde{d}_n^+$ and $\tau^2(v) = \Sigma_i\, d_{ni}^2\, L_{ni}^+(v)$. □

Remark 3.4.4. An alternative proof of (22), using the techniques of Dupač and Hájek (op. cit.), appears in Koul and Staudte, Jr. (1972a). Thus comments like those in Remark 3.4.1 are appropriate here also. □

Next, we turn to the ***weak convergence*** of $\{Z_d\}$ and $\{Z_d^+\}$. These results will be stated without proofs as their proofs are consequences of the results of the previous sections in this chapter.

Therorem 3.4.3. (***Weak convergence of*** Z_d). *Let* $X_{n1}, \ldots, X_{nn}$ *be independent r.v.'s with respective continuous d.f.'s* $F_{n1}, \ldots, F_{nn}$. *With notation as in* (2.2a.33), *assume that* (N1), (N2), (C*) *hold. In addition assume the following:*

(23) *There are measurable functions* $\{\ell_{ni},\ 1 \le i \le n\}$ *on* $[0, 1]$, *such that for all* $0 < k < \infty$,

$$\max_i \sup_{|t-s| \le kn^{-1/2}} n^{1/2} |L_{ni}(t) - L_{ni}(s) - (t-s)\ell_{ni}(s)| = 0$$

Moreover, assume that

(24) $$\limsup_n \sup_{0 \le t \le 1} n^{1/2} |\tilde{d}_n(t)| < \infty,$$

(25) $$\lim_{\delta \to 0} \limsup_n \sup_{|t-s| < \delta} n^{1/2} |\tilde{d}_n(t) - \tilde{d}_n(s)| = 0,$$

(26) $$\liminf_n \sigma^2(t) > 0, \quad 0 < t < 1.$$

Finally, with $K_d(t) := \Sigma_i\, (d_{ni} - \tilde{d}_n(t))\{I(X_{ni} \le H^{-1}(t)) - L_{ni}(t)\}$, *assume that*

(27) $$\begin{aligned} \mathcal{C}(t, s) &= \lim_n \mathrm{Cov}(K_d(t), K_d(s)) \\ &= \lim_n \Sigma_i\, (d_{ni} - \tilde{d}_n(t))(d_{ni} - \tilde{d}_n(s)) L_{ni}(s)(1 - L_{ni}(t)), \end{aligned}$$

exists for all $0 \le s \le t \le 1$.

Then, $Z_d - \mu_d \Rightarrow$ *to a mean zero, covariance* $\mathcal{C}$ *continuous Gaussian process on* $[0, 1]$, *tied down at* 0 *and* 1. □

Remark 3.4.5. In (23), without loss of generality it may be assumed that $n^{-1}\Sigma_i \ell_{ni}(s) = 1$, $0 \le s \le 1$. For, if (23) holds for some $\{\ell_{ni}, 1 \le i \le n\}$, then it also holds for $\{\ell^*_{ni}, 1 \le i \le n\}$, $\ell^*_{ni}(s) := n^{1/2}[L_{ni}(s+n^{-1/2}) - L_{ni}(s)]$, $1 \le i \le n$, $0 \le s \le 1$. Because $n^{-1}\Sigma_i L_{ni}(s) \equiv s$, $n^{-1}\Sigma_i \ell^*_{ni}(s) \equiv 1$. □

Remark 3.4.6. Conditions (C*), (N1) and (24) may be replaced by the condition (B), because, in view of the previous remark,

$$n^{1/2}|\tilde{d}_n(t)| = |n^{-1/2}\Sigma_i d_{ni}\ell_{ni}(t)| \le n^{1/2}\max_i |d_{ni}|, \quad 0 \le t \le 1. \qquad \square$$

Remark 3.4.7. In the case F_{ni} have density f_{ni}, one can choose $\ell_{ni} = f_{ni}(H^{-1})/n^{-1}\Sigma_j f_{nj}(H^{-1})$, $1 \le i \le n$. □

Remark 3.4.8. In the case $F_{ni} \equiv F$, F a continuous and strictly increasing d.f., $L_{ni}(t) \equiv t$, $\ell_{ni}(t) \equiv 1$, so that (C*) and (23) – (26) are trivially satisfied. Moreover, $\mathcal{C}(s,t)$, $= s(1-t)$, $0 \le s \le t \le 1$, so that (27) is satisfied. Thus Theorem 3.4.3 includes Theorem V.3.5.1 of Hájek and Sidák (1967). □

Theorem 3.4.4. (*Weak convergence of* Z_d^+). *Let* $X_{n1}, ..., X_{nn}$ *be independent* r.v.'s *with respective* d.f.'s $F_{n1}, ..., F_{nn}$ *and let* $d_{n1}, ..., d_{nn}$ *be real numbers. Assume that* (N1) *and* (N2) *hold and that the following hold.*

(28) *With* L^+_{dj} *as in* (16),

$$\lim_{\delta\to 0} \limsup_n \sup_{0\le t\le 1-\delta} [L^+_{dj}(t+\delta) - L^+_{dj}(t)] = 0, \qquad j = 1, 2.$$

(29) *There are measurable functions* ℓ^+_{ij}, $1 \le i \le n$, $j = 1, 2$ *on* [0, 1] *such that for any* $0 < k < \infty$,

$$\max_i \sup_{|t-s|\le kn^{-1/2}} n^{1/2}|L^+_{ij}(t) - L^+_{ij}(s) - (t-s)\ell^+_{ij}(s)| = o(1).$$

(30) *With* $\tilde{d}^+_n$ *as in* (20),

$$\limsup_n \sup_{0\le t\le 1} n^{1/2}|\tilde{d}^+_n(t)| < \infty,$$

$$(31) \qquad \lim_{\delta\to 0} \limsup_n \sup_{|t-s|<\delta} n^{1/2}|\tilde{d}^+_n(t) - \tilde{d}^+_n(s)| = 0.$$

(32) *With* τ^2 *as in* (20),

$$\liminf_n \tau^2(t) > 0, \qquad 0 < t < 1.$$

(33) *With* $K_d^+(t)$ *as in the proof of Theorem* 3.4.2,

$$\lim_n \operatorname{Cov}(K_d^+(s), K_d^+(t)) = \mathcal{C}^+(s,t) \text{ exists}, \qquad 0 \le s \le t \le 1.$$

Then, $Z_d^+ - \mu_d^+ \Rightarrow$ *to a continuous mean zero covariance* $\mathcal{C}^+$ *Gaussian process, tied down at* 0. □

Remark 3.4.9. Remarks 3.4.5 through 3.4.7 are applicable here also, with appropriate modifications. □

Remark 3.4.10. Suppose that $F_{ni} \equiv F$, F continuous, and $d_{ni} \equiv n^{-1/2}$. Then

$$\sup_{0 \le t \le 1} |Z_d^+(t) - \mu_d^+(t)| = \sup_{0<x<\infty} n^{1/2} |\{H_n(x) - H_n(0)\} - \{H_n(0) - H_n(-x)\} - \{F(x) - F(0)\} - \{F(0) - F(-x)\}|$$

which is precisely the statistic τ_n^* proposed by Smirnov (1947) to test the hypothesis of symmetry about F. Smirnov considered only the null distribution. Theorem 3.4.4 allows one to study its asymptotic distribution under fairly general independent alternatives.

If $\{d_{ni}\}$ are arbitrary, subject to (N1) and (N2), then $\sup\{|Z_d^+(t) - \mu_d^+(t)|;\ 0 \le t \le 1\}$ may be considered a generalized Smirnov statistic for testing the hypothesis of symmetry. □□

CHAPTER 4

M, R AND SOME SCALE ESTIMATORS

4.1. INTRODUCTION

In the last three decades statistics has seen the emergence and consolidation of many competitors of the Least Square estimator of β of (1.1.1). The most prominent are the so-called M- and R- estimators. The class of M-estimators was introduced by Huber (1973) and its computational aspects and some robustness properties are available in Huber (1981). The class of R-estimators is based on the ideas of Hodges and Lehmann (1963) and has been developed by Adichie (1967), Jurečková (1971) and Jaeckel (1972).

One of the attractive features of these estimators is that they are robust against certain outliers in errors. All of these estimators are translation invariant, whereas only R-estimators are scale invariant.

Our purpose here is to illustrate the usefulness of the results of Chapter 2 in deriving the asymptotic distributions of these estimators under a fairly general class of heteroscedastic errors. Section 4.2a gives the asymptotic distributions of M-estimators while those of R-estimators are given in Section 4.4. Among other things, the results obtained enable one to study their qualitative robustness against an array of non-identical error d.f.'s converging to a fixed error d.f. The sufficient conditions given here are fairly general for the underlying score functions and the design variables.

Efron (1979) introduced a general resampling procedure, called the bootstrap, for estimating the distribution of a pivotal statistic. Singh (1981) showed that the bootstrap estimate B_n is *second order accurate*, i.e., provides more accurate approximation to the sampling distribution G_n of the standardized sample mean than the usual normal approximation in the sense that $\sup\{|G_n(x) - B_n(x)|;\ x\in\mathbb{R}\}$ tends to zero at a faster rate than that of the square-root of n. This kind of result holds more generally as noted by Babu and Singh (1983, 1984).

Section 4.2b discusses similar results pertaining to a class of M-estimators of β when the errors in (1.1.1) are i.i.d.. It is noted that Shorack's (1982) modified bootstrap estimator and the one obtained by resampling the residuals according to a w.e.p. are second order accurate.

In an attempt to make M-estimators scale invariant one often needs a preliminary robust scale estimator. Two such estimators are the *MAD* (median of absolute deviations of residuals) and the *MASD* (median of absolute symmetrized deviations of residuals). The asymptotic distributions of these estimators under heteroscedastic errors appear in Section 4.3.

In carrying out the analysis of variance of an experimental design or a linear model based on ranks one needs an estimator of the asymptotic variance of certain rank statistics, see, e.g., Hettmansperger (1984). These variances involve the functional. $Q(f) = \int f\, d\varphi(F)$ where φ is a known

function, F a common error d.f. having a density f. Some estimators of Q(f) under (1.1.1) are presented in Section 4.5. Again, the results of Chapter 2 are found useful in proving their consistency.

4.2. M-ESTIMATORS

4.2a. First Order Approximations: Asymptotic Normality

This subsection contains the asymptotic distributions of M-estimators of $\boldsymbol{\beta}$ when the errors in (1.1.1) are heteroscedastic. The following subsection 4.2b gives some results on the bootstrap approximations to these distributions.

Let the model (1.1.1) hold. Let ψ be a nondecreasing function from $\mathbb{R}$ to $\mathbb{R}$. The corresponding M-estimator $\hat{\boldsymbol{\Delta}}$ of $\boldsymbol{\beta}$ is defined to be a zero of the M-score $\int\psi(y)\, \mathbf{V}(dy, \mathbf{t})$, where $\mathbf{V}$ is defined at (1.1.2). Our objective is to investigate the asymptotic behavior of $\mathbf{A}^{-1}(\hat{\boldsymbol{\Delta}} - \boldsymbol{\beta})$ when the errors in (1.1.1) are heteroscedastic. Our method is still the usual one, v.i.z., to obtain the expansion of the M-score uniformly in $\mathbf{t} \in \{\mathbf{t};\ \|\mathbf{A}^{-1}(\mathbf{t}-\boldsymbol{\beta})\| \le B\}$, $0 < B < \infty$, to observe that there is a zero of the M-score, $\hat{\boldsymbol{\Delta}}$, in this set and then to apply this expansion to obtain the approximation for $\mathbf{A}^{-1}(\hat{\boldsymbol{\Delta}} - \boldsymbol{\beta})$ in terms of the given M-score at the true $\boldsymbol{\beta}$. To make all this precise, we need to standardize the M-score. For that reason we need some more notation. Let

$$
\begin{aligned}
(1)\qquad & \Lambda^*(y) := \mathrm{diag}(f_{n1}(y), \ldots, f_{nn}(y)), && y \in \mathbb{R},\\
& \mathcal{C} := \mathbf{A}\mathbf{X}'\int \Lambda^*(y)\, d\psi(y)\, \mathbf{X}\mathbf{A}, &&\\
& \mathbf{T}(\psi, \mathbf{t}) := -\mathcal{C}^{-1}\mathbf{A}\int \psi(y)\mathbf{V}(dy, \mathbf{t}), &&\\
& \overline{\mathbf{T}}(\psi, \mathbf{t}) := \mathbf{A}^{-1}(\mathbf{t} - \boldsymbol{\beta}) - \mathbf{T}(\psi, \boldsymbol{\beta}), && \mathbf{t} \in \mathbb{R}^p.
\end{aligned}
$$

An approximation to $\hat{\boldsymbol{\Delta}}$ is given by the zero $\overline{\boldsymbol{\Delta}}$ of $\overline{\mathbf{T}}(\psi, \mathbf{t})$, v.i.z.,

$$
(2)\qquad \mathbf{A}^{-1}(\overline{\boldsymbol{\Delta}} - \boldsymbol{\beta}) = \mathbf{T}(\psi, \boldsymbol{\beta}).
$$

A basic result needed to make this precise is the a.u.l. of $\mathbf{T}(\psi, \mathbf{t})$ in $\mathbf{A}^{-1}(\mathbf{t} - \boldsymbol{\beta})$. Often such a result is obtained under some smooth conditions on ψ and under i.i.d. errors. Theorem 4.2a.1 below gives such a result for a general nondecreasing right continuous bounded ψ and for fairly general independent heteroscedastic errors.

Theorem 4.2a.1. *Let* $\{(\mathbf{x}_{ni}', Y_{ni}),\ 1 \le i \le n\}$, β, $\{F_{ni},\ 1 \le i \le n\}$ *be as in the model* (1.1.1) *satisfying all conditions of Theorem* 2.3.3. *In addition, assume the following:*

(3) $\quad \psi \in \Psi := \{\psi\colon \mathbb{R}$ *to* $\mathbb{R}$, $\psi \in \mathbb{DI}(\mathbb{R})$, *bounded with* $|\psi|_{tv} \le k < \infty\}$.

(4) $$\limsup_n \|C^{-1}\|_\infty < \infty.$$

Then, $\forall\ 0 < B < \infty$,

(5) $$\sup_{\psi, \mathbf{u}} \|\mathbf{T}(\psi, \beta + \mathbf{A}\mathbf{u}) - \overline{\mathbf{T}}(\psi, \beta + \mathbf{A}\mathbf{u})\| = o_p(1).$$

where the supremum is taken over all $\psi \in \Psi$ *and* $\|\mathbf{u}\| \le B$.

Proof: Rewrite, after integration by parts,

$$\mathbf{T}(\psi, \mathbf{t}) - \overline{\mathbf{T}}(\psi, \mathbf{t}) = \int C^{-1}\mathbf{A}\,[\mathbf{V}(y, \mathbf{t}) - \mathbf{V}(y, \beta) - \Gamma_1(H(y))\mathbf{A}^{-1}(\mathbf{t} - \beta)]d\psi(y).$$

Now (5) readily follows from this and (2.3.37). □

In order to use this theorem, we must be able to argue that $\|\mathbf{A}^{-1}(\hat{\Delta} - \beta)\| = O_p(1)$. To that effect, define

$$\mu_i := E\,\psi(e_i), \quad \tau_i^2 = \mathrm{Var}\,\psi(e_i), \quad 1 \le i \le n,$$

$$\mathbf{b}_n := E\,\mathbf{T}(\psi, \beta) = -C^{-1}\mathbf{A}\,\Sigma_i\,\mathbf{x}_i\,\mu_i$$

and observe that

$$E\|\mathbf{A}^{-1}(\bar{\Delta} - \beta) - \mathbf{b}_n\|^2 = C^{-1}\,\Sigma_i\,\mathbf{x}_i'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i\,\tau_i^2\,C^{-1} = O(1),$$

by (3), (4) and the fact that $\Sigma_i\,\mathbf{x}_i'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i \equiv p < \infty$. Hence by the Markov inequality, $\forall\ \epsilon > 0\ \exists\ 0 < K_\epsilon < \infty\ \ni$

$$P(\|\mathbf{A}^{-1}(\bar{\Delta} - \beta) - \mathbf{b}_n\| \le K_\epsilon) \ge 1 - \epsilon, \text{ for all } n \ge 1.$$

Thus, assuming that

(6) $$\Sigma_i\,\mathbf{x}_i\,\mu_i = \mathbf{0},$$

and arguing via Brouwer's fixed point theorem as in Huber (1981,p 169), one concludes, in view (5), that $\forall\ \epsilon > 0\ \exists\ N_\epsilon$ and K_ϵ such that

$$\text{(7)} \qquad P(\|A^{-1}(\hat{\Delta} - \beta)\| \le K\epsilon) \ge 1 - \epsilon, \quad \forall\ n \ge N\epsilon.$$

Now, a routine application of (5) enables one to conclude that

$$0 = T(\psi, \hat{\Delta}) = A^{-1}(\hat{\Delta} - \beta) - T(\psi, \beta) + o_p(1),$$

i.e.,

$$\text{(8)} \qquad A^{-1}(\hat{\Delta} - \beta) = T(\psi, \beta) + o_p(1).$$

Note that, under (6), with $T_0(\psi, \beta) = C\, T(\psi, \beta)$,

$$E\, T_0(\psi, \beta)\, T_0'(\psi, \beta) = A\, \Sigma_i\, x_i' x_i\, \tau_i^2\, A = AX'\mathcal{T} XA$$

where $\mathcal{T} = \text{diag}\,(\tau_1^2, \ldots, \tau_n^2)$. Moreover, for any $\lambda \in \mathbb{R}^p$,

$$\lambda' T_0(\psi, \beta) = \Sigma_i \sum_{j=1}^{p} \lambda_j d_{ij} \psi(e_i) = \Sigma_i\, \lambda' A\, x_i'\, \psi(e_i)$$

where $\{d_{ij}\}$ are as in (2.3.32). In view of (2.3.33) and (2.3.34), (NX) and (6) imply that $\lambda' T_0(\psi, \beta)$ is asymptotically normally distributed with mean 0 and the asymptotic variance $\lambda' AX'\mathcal{T} XA\, \lambda$. Thus by the Cramer-Wold device [Theorem 7.7, p 49, Billingsley (1968)], (4) and (8),

$$\text{(9)} \qquad \Sigma^{-1/2} A^{-1}(\hat{\Delta} - \beta) \underset{d}{\rightarrow} N(0, I_{p\times p}),\ \Sigma := C^{-1} AX'\mathcal{T} XAC^{-1}.$$

We summarize the above discussion as a

Proposition 4.2a.1. *Suppose that the d.f.'s* $\{F_{ni}\}$ *of the errors and the design matrix* X *of* (1.1.1) *satisfy* (4), (6) *and the assumptions of Theorem* 2.3.3 *including that* H *is strictly increasing for each* $n \ge 1$. *Then* (9) *holds.* □

Now, consider the case of the *i.i.d. errors* in (1.1.1) with $F_{ni} \equiv F$. Then,

$$\text{(10)} \qquad \tau_i^2 = \int \psi^2\, dF - \left(\int \psi dF\right)^2 = \tau^2, \text{ (say)}, \qquad 1 \le i \le n,$$
$$C = \left(\int f\, d\psi\right) I_{p\times p}, \qquad \Sigma = \left(\int f\, d\psi\right)^{-2} \tau^2\, I_{p\times p}.$$

Consequently (4) is equivalent to requiring $\int f\, d\psi > 0$. Next, observe that (6) becomes

$$\text{(6*)} \qquad \Sigma_i\, x_i \int \psi\, dF = 0.$$

Obviously, this is satisfied if either $\Sigma_i \mathbf{x}_i = \mathbf{0}$, i.e., if $\mathbf{X}$ is a centered design matrix or if $\int \psi \, dF = 0$, the often assumed condition. The former excludes the possibility of the presence of the location parameter in (1.1.1). Thus to summarize, we have

Proposition 4.2a.2. *Suppose that in* (1.1.1), $F_{ni} \equiv F$. *In addition, assume that* X *and* F *satisfy* (**NX**), (**F1**), (**F2**), (6*) *and that* $\int f \, d\psi > 0$. *Then,*

$$\mathbf{A}^{-1}(\hat{\Delta} - \boldsymbol{\beta}) \underset{d}{\longrightarrow} N_p(0, \tau^2/(\int f \, d\psi)^2 \, \mathbf{I}_{p\times p}). \qquad \square$$

Condition (6*) suggests another way of defining M-estimators of $\boldsymbol{\beta}$ in (1.1.1) in the case of the *i.i.d. errors.* Let

$$\text{(11)} \qquad \bar{x}_{nj} := n^{-1} \sum_{i=1}^{n} x_{nij}, \quad 1 \le j \le p; \qquad \bar{\mathbf{x}}_n' := (\bar{x}_{n1}, \dots, \bar{x}_{np}),$$

$$\bar{\mathbf{X}}' := [\,\bar{\mathbf{x}}_n, \dots, \bar{\mathbf{x}}_n]_{n\times p}, \qquad \mathbf{X}_c := \mathbf{X} - \bar{\mathbf{X}}.$$

Assume

(**NX1**)
$$(\mathbf{X}_c'\mathbf{X}_c)^{-1} \quad \text{exists for all } n \ge p,$$
$$\max_i \mathbf{x}_{ni}' (\mathbf{X}_c'\mathbf{X}_c)^{-1} \mathbf{x}_{ni} = o(1).$$

Let

$$\text{(12)} \qquad \mathbf{T}^*(\psi, \mathbf{t}) := \mathbf{A}_1 \sum_i (\mathbf{x}_{ni} - \bar{\mathbf{x}}_n)\, \psi(Y_i - \mathbf{x}_{ni}'\mathbf{t}), \qquad \mathbf{t} \in \mathbb{R}^p,$$
$$\mathbf{A}_1 := (\mathbf{X}_c'\mathbf{X}_c)^{-1/2}.$$

Define an M-estimator Δ^* as a solution $\mathbf{t}$ of

$$\text{(13)} \qquad \mathbf{T}^*(\psi, \mathbf{t}) = \mathbf{0}.$$

Apply Corollary 2.3.1 p times, j^{th} time with $d_{ni} = i^{th}$ element of the j^{th} column of $\mathbf{X}_c\mathbf{A}_1$, $1 \le i \le n$, $1 \le j \le p$, to conclude an analogue of (5) above, v.i.z.,

$$\text{(5*)} \qquad \sup_{\psi, \|\mathbf{A}_1^{-1}(\mathbf{t}-\boldsymbol{\beta})\| \le B} \|\mathbf{T}^*(\psi, \mathbf{t}) - \bar{\mathbf{T}}^*(\psi, \mathbf{t})\| = o_p(1)$$

where

$$\bar{\mathbf{T}}^*(\psi, \mathbf{t}) := \mathbf{A}_1^{-1}(\mathbf{t} - \boldsymbol{\beta}) - (\int f \, d\psi)^{-1} \, \mathbf{T}^*(\psi, \boldsymbol{\beta}).$$

The proof of (5*) is exactly similar to that of (5) with appropriate modifications of replacing $\mathbf{X}$ by $\mathbf{X}_c$ and $\mathbf{A}$ by $\mathbf{A}_1$ and using $F_{ni} \equiv F$ in the discussion there.

Now, clearly, $F_{ni} \equiv F$ implies that $E\, \mathbf{T}^*(\psi, \boldsymbol{\beta}) = \mathbf{0}$,

$$E\|\mathbf{T}^*(\psi, \boldsymbol{\beta})\|^2 = \Sigma_i (\mathbf{x}_i - \bar{\mathbf{x}})' \mathbf{A}_1' \mathbf{A}_1 (\mathbf{x}_i - \bar{\mathbf{x}})\, \tau^2 = O(1).$$

Hence, $\|\mathbf{T}^*(\psi, \boldsymbol{\beta})\| = O_p(1)$. If $\bar{\boldsymbol{\Delta}}^*$ is zero of $\bar{\mathbf{T}}^*(\psi, .)$, then

$$\mathbf{A}_1^{-1}(\bar{\boldsymbol{\Delta}}^* - \boldsymbol{\beta}) = \left(\int f\, d\psi\right)^{-1} \mathbf{T}^*(\psi, \boldsymbol{\beta}).$$

Argue, as for (7), (8) and (9) to conclude the following

Proposition 4.2a.3. *Suppose that in* (1.1.1), $F_{ni} \equiv F$. *In addition, assume that* $\mathbf{X}$ *and* F *satisfy* **(NX1)**, **(F1)** *and* **(F2)**. *Then,*

$$\mathbf{A}_1^{-1}(\boldsymbol{\Delta}^* - \boldsymbol{\beta}) \underset{d}{\rightarrow} N_p(\mathbf{0}, \tau^2/(\textstyle\int f\, d\psi)^2\, \mathbf{I}_{p\times p}),$$

where $\boldsymbol{\Delta}^*$ *is as in* (13). □

Remark 4.2a.2. Note that the Proposition 4.2a.3 does not require the condition $\int \psi\, dF = 0$. An advantage of this is that $\boldsymbol{\Delta}^*$ can be used as a preliminary estimator when constructing adaptive estimators of $\boldsymbol{\beta}$. An *adaptive* estimator is one that achieves the Hájek – Le Cam (Hájek 1972, Le Cam 1972) lower bound over a large class of error distributions. Often, a minimal condition required to construct an adaptive estimator of $\boldsymbol{\beta}$ is that F have finite Fisher information, i.e., that F satisfy (3.2.a) of Theorem 3.2.3. See, e.g., Bickel (1982), Fabian and Hannan (1982) and Koul and Susarla (1983). Recall, from Remark 3.2.2, that this implies **(F1)**.

On the other hand, the condition **(NX1)** does not allow for any location term in the linear regression model. □

So far we have been dealing with the linear regression model with known scale. Now consider the model (2.3.38) where γ is an unknown scale parameter. Let s be an $n^{1/2}$ – consistent estimator of γ, i.e.,

$$(14) \qquad |n^{1/2}(s - \gamma)\gamma^{-1}| = O_p(1).$$

Define an M-estimator $\hat{\boldsymbol{\Delta}}_1$ of $\boldsymbol{\beta}$ as a solution $\mathbf{t}$ of

$$(15) \qquad \Sigma\, \mathbf{x}_i\, \psi((Y_i - \mathbf{x}_i'\mathbf{t})s^{-1}) = \mathbf{0} \quad \text{or} \quad \int \psi(y)\, \mathbf{V}(sdy, \mathbf{t}) = \mathbf{0}.$$

To keep exposition simple, now we shall not exhibit ψ in some of the functions defined below. Define, for an $\alpha > 0$, $\mathbf{t} \in \mathbb{R}^p$,

$$(16) \quad \mathbf{S}(\alpha, \mathbf{t}) := -\mathbf{A} \int \psi(y)\, \mathbf{V}(\alpha dy, \mathbf{t}),$$

$$\overline{\mathbf{S}}(\alpha, \mathbf{t}) := \mathbf{A}^{-1}(\mathbf{t}-\boldsymbol{\beta})\gamma^{-1} + \mathcal{C}^{-1}\mathcal{C}_1\, n^{1/2}(\alpha-\gamma)\gamma^{-1} - \mathcal{C}^{-1}\,\mathbf{S}(\gamma, \boldsymbol{\beta}),$$

where

$$\mathcal{C}_1 := n^{-1/2}\mathbf{A}\mathbf{X}' \int y f(y)\, d\psi(y), \quad f'(y) := (f_1(y), \ldots, f_n(y)),$$

and where $\mathcal{C}$ is as in (1) above. Note that by (**NX**), (**F1**), (**F3**), and (3),

$$(17) \qquad \|\mathcal{C}_1\| = O(1).$$

The following theorem is a direct consequence of Theorem 2.3.4. In it $N_1 := \{(\alpha, \mathbf{t})\colon \alpha > 0,\ \mathbf{t}\in\mathbb{R}^p,\ \|\mathbf{A}^{-1}(\mathbf{t}-\boldsymbol{\beta})\| \le B\gamma,\ |n^{1/2}(\alpha-\gamma)| \le b\gamma\}$, $0<b, B<\infty$.

Theorem 4.2a.2. *Let* $\{(\mathbf{x}'_{ni}, Y_{ni}),\ 1 \le i \le n\}$, $\boldsymbol{\beta}$, γ, $\{F_{ni},\ 1 \le i \le n\}$ *be as in* (2.3.38) *satisfying all the conditions of Theorem* 2.3.4. *Moreover, assume* (3) *and* (4) *hold. Then, for every* $0 < b, B < \infty$ *fixed,*

$$(18) \qquad \sup \|\mathbf{S}(\alpha, \mathbf{t}) - \overline{\mathbf{S}}(\alpha, \mathbf{t})\| = o_p(1).$$

where the supremum is taken over all $\psi \in \Psi$, *and* $(\alpha, \mathbf{t}')' \in N_1$. □

Now argue as in the proof of the Proposition 4.2a.1 to conclude

Proposition 4.2a.3. *Suppose that the design matrix* $\mathbf{X}$ *and* d.f.'s $\{F_{ni}\}$ *of* $\{\epsilon_{ni}\}$ *in* (2.3.38) *satisfy* (5), (6) *and the assumptions of Theorem* 2.3.4 *including that* H *is strictly increasing for each* $n \ge 1$. *In addition assume that there exists an estimate* s *of* γ *satisfying* (14). *Then*

$$(19) \quad \mathbf{A}^{-1}(\hat{\boldsymbol{\Delta}}_1 - \boldsymbol{\beta})\gamma^{-1} = \mathcal{C}^{-1}\mathbf{S}(\gamma, \boldsymbol{\beta}) - \mathcal{C}^{-1}\mathcal{C}_1 n^{1/2}(s-\gamma)\,\gamma^{-1} + o_p(1),$$

where $\hat{\boldsymbol{\Delta}}_1$ *now is a solution of* (15). □

Remark 4.2a.3. In (6), F_i is now the d.f. of ϵ_i, and not of $\gamma\epsilon_i$, $1 \le i \le n$. □

Remark 4.2a.4. ***Effect of symmetry on*** $\hat{\boldsymbol{\Delta}}_1$. As is clear from (19), in general the asymptotic distribution of $\hat{\boldsymbol{\Delta}}_1$ depends on s. However, suppose that

$$(20) \qquad d\psi(y) = -\,d\psi(-y), \quad f_i(y) \equiv f_i(-y), \quad 1 \le i \le n, \quad -\infty < y < +\infty.$$

Then $\int y\, f_i(y)\, d\psi(y) = 0$, $1 \le i \le n$, and, from (16), $\mathcal{C}_1 = \mathbf{0}$. Consequently,

in this case,

$$A^{-1}(\hat{\Delta}_1 - \beta) = \gamma C^{-1} S(\gamma, \beta) + o_p(1).$$

Hence, with Σ as in (9), we obtain

$$\Sigma^{-1/2} A^{-1}(\hat{\Delta}_1 - \beta)\, \gamma^{-1} \xrightarrow[d]{} N_p(0, I_{p\times p}). \tag{21}$$

Note that this asymptotic distribution differs from that of (9) only by the presence of γ^{-1}. In other words, in the case of symmetric errors $\{\epsilon_i\}$ and the skew symmetric score functions $\{\psi\}$, the asymptotic distribution of M-estimator of β of (2.3.38) with a preliminary $n^{1/2}$-consistent estimator of the scale parameter is the same as that of $\gamma^{-1}\times$M-estimator of β of (1.1.1). □

4.2b. Bootstrap Approximations

Before discussing the specific bootstrap approximations we shall describe the concept of Efron's bootstrap a bit more generally in the one sample setup.

Let $\xi_1, \xi_2, \ldots, \xi_n$ be n i.i.d. G r.v.'s, G_n be their empirical d.f. and $T_n = T_n(\xi_n, G)$ be a function of $\xi_n' := (\xi_1, \xi_2, \ldots, \xi_n)$ and G such that $T_n(\xi_n, G)$ is a r.v. for every G. Let $\zeta_1, \zeta_2, \ldots, \zeta_n$ denote i.i.d. G_n r.v.'s and $\zeta_n' := (\zeta_1, \zeta_2, \ldots, \zeta_n)$. The *bootstrap* d.f. B_n of $T_n(\xi_n, G)$ is the d.f. of $T_n(\zeta_n, G_n)$ under G_n. Efron (1979) showed, via numerical studies, that in several examples B_n provides better approximation to the d.f. Γ_n of $T_n(\xi_n, G)$ under G than the normal approximation. Singh (1981) substantiated this observation by proving that in the case of the standardized sample mean the bootstrap estimate B_n is *second order accurate*, i.e.,

$$\sup\{|\Gamma_n(x) - B_n(x)|;\, x \in \mathbb{R}\} = o(n^{-1/2}), \text{ a.s..} \tag{1}$$

Recall that the Edgeworth expansion or the Berry-Esseen bound gives that

$$\sup\{|\Gamma_n(x) - \Phi(x)|;\, x \in \mathbb{R}\} = O(n^{-1/2}),$$

where Φ is the d.f. of a N(0, 1) r.v. See, e.g., Feller (1966. Ch. XVI). Babu and Singh (1983, 1984), among others, pointed out that this phenomenon is shared by a large class of statistics. For further reading on bootstrapping we refer the reader to Efron (1982).

We now turn to the problem of bootstrapping M-estimators in a linear regression model. For the sake of clarity we shall restrict our attention to a *simple* linear regression model only. Our main purpose is to show how a certain weighted empirical sampling distribution naturally helps to overcome

some inherent difficulties in defining bootstrap M-estimators. What follows is based on the work of Lahiri (1989). No proofs will be given as they involve intricate technicalities of the Edgeworth expansion for independent non-identically distributed r.v.'s.

Accordingly, assume that $\{e_i, i \geq 1\}$ are i.i.d F r.v.'s, $\{x_{ni}, i \geq 1\}$ are the known design points, $\{Y_{ni}, i \geq 1\}$ are observable r.v.'s such that for a $\beta \in \mathbb{R}$,

$$Y_{ni} = x_{ni}\beta + e_i, \qquad i \geq 1. \tag{2}$$

The score function ψ is assumed to satisfy

$$\int \psi \, dF = 0. \tag{3}$$

Let $\hat{\Delta}_n$ be an M-estimator obtained as a solution t of

$$\sum_{i=1}^{n} x_{ni}\, \psi(Y_{ni} - x_{ni}t) = 0 \tag{4}$$

and F_n be an estimator of F based on the residuals $\hat{e}_{ni} := Y_{ni} - x_{ni}\hat{\Delta}_n$, $1 \leq i \leq n$. Let $\{e^*_{ni}, 1 \leq i \leq n\}$ be i.i.d. F_n r.v.'s and define

$$Y^*_{ni} = x_{ni}\hat{\Delta}_n + e^*_{ni}, \qquad 1 \leq i \leq n. \tag{5}$$

The bootstrap M-estimator Δ^*_n is defined to be a solution t of

$$\sum_{i=1}^{n} x_{ni}\, \psi(Y^*_{ni} - x_{ni}t) = 0. \tag{6}$$

Recall, from the previous section, that in general (3) ensures the absence of the asymptotic bias in $\hat{\Delta}_n$. Analogously, to ensure the absence of the asymptotic bias in Δ^*_n, we need to have F_n such that

$$\int \psi \, dF_n = E_n\, \psi(e^*_{n1}) = 0, \tag{7}$$

where E_n is the expectation under F_n. In general, the choice of F_n that will satisfy (7) and at the same time be a reasonable estimator of F depends heavily on the forms of ψ and F. When bootstrapping the least square estimator of β, i.e., when $\psi(x) \equiv x$, Freedman (1981) ensures (7) by choosing F_n to be the empirical d.f. of the centered residuals $\{\hat{e}_{ni} - \hat{e}_{n\cdot}, 1 \leq i \leq n\}$, where $\hat{e}_{n\cdot} := n^{-1}\sum_{j=1}^{n} \hat{e}_{nj}$. In fact, he shows that if one does not center the residuals, the bootstrap distribution of the least squares estimator does not approximate the corresponding original distribution.

Clearly, the ordinary empirical d.f. $\hat{H}_n$ of the residuals $\{\hat{e}_{ni};\ 1\le i\le n\}$ does not ensure the validity of (7) for general designs and a general ψ. We are thus forced to look at appropriate modifications of the usual bootstrap. Here we describe two modifications. One chooses the resampling distribution appropriately and the other modifies the defining equation (6) *a la* Shorack (1982). Both provide the second order correct approximations to the distribution of standardized $\hat{\Delta}_n$.

Weighted Empirical Bootstrap:

Assume that the design points $\{x_{ni}\}$ are either all non–negative or all non–positive. Let $\omega_x := \sum_{i=1}^n |x_{ni}|$ be positive and define

$$F_{1n}(y) := \omega_x^{-1}\sum_{i=1}^n |x_{ni}|\, I(\hat{e}_{ni} \le y), \qquad y\in\mathbb{R}. \tag{8}$$

Take the resampling distribution F_n to be F_{1n}. Then, clearly,

$$E_{1n}\,\psi(e_{n1}^*) = \omega_x^{-1}\sum_{i=1}^n |x_{ni}|\,\psi(\hat{e}_{ni}) = \mathrm{sign}(x_1)\,\omega_x^{-1}\sum_{i=1}^n x_{ni}\psi(Y_{ni}-x_{ni}\hat{\Delta}) = 0,$$

by the definition of $\hat{\Delta}_n$. That is, F_{1n} satisfies (7) for any ψ.

Modified Scores Bootstrap:

Let F_n be any resampling distribution based on the residuals. Define the bootstrap estimator Δ_{ns} to be a solution t of the equation

$$\sum_{i=1}^n x_{ni}\,[\psi(Y_{ni}^* - x_{ni}t) - E_n\,\psi(e_{n1}^*)] = 0. \tag{9}$$

In other words the score function is now *a priori* centered under F_n and hence (7) holds for any F_n and any ψ.

We now describe the the second order correctness of these procedures. To that effect we need some more notation and assumptions. To begin with let $\tau_x^2 := \sum_{i=1}^n x_{ni}^2$ and define

$$m_x := \max\{|x_{ni}|;\ 1\le i\le n\},\quad b_{1x} := \sum_{i=1}^n x_{ni}^3/\tau_x^3,\quad b_x := \sum_{i=1}^n |x_{ni}^3|/\tau_x^3.$$

For a d.f. F and any sampling d.f. F_n, define

$$\gamma(x) := E\psi(e_1 - x), \qquad \varpi(x) = \sigma^2(x) := E\{\psi(e_1-x) - \gamma(x)\}^2,$$

$$\mathscr{w}_1(x) := E\{\psi(e_1 - x) - \gamma(x)\}^3, \qquad x \in \mathbb{R}.$$

$$\gamma_n(x) := E_n\psi(\overset{*}{e}_{n1} - x), \quad \mathscr{w}_n(x) = \sigma_n^2(x) := E_n\{\psi(\overset{*}{e}_{n1} - x) - \gamma_n(x)\}^2,$$

$$\mathscr{w}_{1n}(x) := E_n\{\psi(\overset{*}{e}_{n1} - x) - \gamma_n(x)\}^3, \qquad x \in \mathbb{R}.$$

$$A_n(c) := \{i\colon\ 1 \le i \le n,\ |x_{ni}| > c\tau_x\, b_x\}, \quad \kappa_n(c) := \#A_n(c), \quad c > 0.$$

For any real valued function g on $\mathbb{R}$, let $\dot{g}$, $\ddot{g}$ denote its first and second derivatives at 0 whenever they exist, respectively. Also, write γ_n, $\mathscr{w}_n$ etc. for $\gamma_n(0)$, $\mathscr{w}_n(0)$, etc. Finally, let $\alpha := -\dot{\gamma}/\sigma$, $\alpha_n := -\dot{\gamma}_n/\sigma_n$, and, define for $x \in \mathbb{R}$, $H_2(x) := x^2 - 1$, and

$$\mathscr{G}_n(x) := \Phi(x) - b_{1x}\,[\{\ddot{\gamma}_n/\sigma_n - \dot{\gamma}_n\ \dot{\mathscr{w}}_n/\sigma_n^3\}(x^2/2\alpha_n^2) + (\mathscr{w}_{1n}/6\sigma_n^3)\, H_2(x)]\ \varphi(x).$$

In the following theorems, a.s. means for almost all sequences $\{e_i;\ i \ge 1\}$ of i.i.d. F r.v.'s.

Theorem 4.2b.1. *Let the model (2) hold. In addition assume that ψ has uniformly continuous bounded second derivative and that the following hold:*

(a) $\tau_x^2 \longrightarrow \infty$. (b) $\alpha > 0$.

(c) *There exists a constant* $0 < c < 1$, *such that* $\ell n\ \tau_x = o(\kappa_n(c))$.

(d) $m_x\ \ell n\ \tau_x = o(\tau_x)$.

(e) *There exist constants* $\theta > 0$, $\delta > 0$ *and* $q < 1$ *such that*

$$\sup[\,|E\exp\{it\psi(e_1 - x)\}|\colon\ |x| < \delta,\ |t| > \theta] < q.$$

(f) $\forall\ \lambda > 0,\ \sum_{n=1}^{\infty} \exp(-\lambda\omega_x^2/\tau_x^2) < \infty$.

Then, with $\overset{*}{\Delta}_n$ *defined as a solution of* (6) *with* $F_n = F_{1n}$,

$$\sup_y |P_{1n}(\alpha_n\tau_x(\overset{*}{\Delta}_n - \hat{\Delta}_n) \le y) - \mathscr{G}_n(y)| = o(m_x/\tau_x),$$

$$\sup_y |P_{1n}(\alpha\tau_x(\hat{\Delta}_n - \beta) \le y) - P_{1n}(\tau_x(\overset{*}{\Delta}_n - \hat{\Delta}_n) \le y)| = o(m_x/\tau_x), \text{ a.s.},$$

where P_{1n} *denotes the bootstrap probability under* F_{1n}, *and where the supremum is over* $y \in \mathbb{R}$. □

Next we state the analogous result for Δ_{ns}.

Theorem 4.2b.2. *Suppose that all of the hypotheses of Theorem* 4.2b.1 *except* (f) *hold and that* Δ_{ns} *is defined as a solution of* (9) *with* $F_n = \hat{H}_n$, *the*

ordinary empirical of the residuals. Then,

$$\sup_y |\hat{P}_n(\alpha_n \tau_x(\Delta_{ns} - \hat{\Delta}_n) \leq y) - \mathscr{F}_n(y)| = o(m_x/\tau_x),$$

$$\sup_y |\hat{P}_n(\alpha\tau_x(\hat{\Delta}_n - \beta) \leq y) - \hat{P}_n(\tau_x(\Delta_{ns} - \hat{\Delta}_n) \leq y)| = o(m_x/\tau_x), \text{ a.s.},$$

where $\hat{P}_n$ *denotes the bootstrap probability under* $\hat{H}_n$. □

The proofs of these theorems appear in Lahiri (1989) where he also discusses analogous results for a non–smooth ψ. In this case he chooses the sampling distribution to be a smooth estimator obtained from the kernel type density estimator. Lahiri (1991) gives extensions of the above theorems to multiple linear regression models.

Here we briefly comment about the assumptions (a) – (f). As is seen from the previous section, (a) and (b) are minimally required for the asmyptotic normality of M–estimators. Assumptions (c), (e) and (f) are required to carry out the Edgeworth expansions while (d) is slightly stronger than Noether's condition (NX) applied to (2). In particular, $x_i \equiv 1$ and $x_i \equiv i$ satisfy (a), (c), (d) and (f).

A sufficient condition for (e) to hold is that F have a positive density and ψ have a continuous positive derivative on an open interval in $\mathbb{R}$.

4.3. DISTRIBUTION OF SOME SCALE ESTIMATORS

Here we shall now discuss some robust scale estimators.

Definitions. An estimator $\hat{\beta}(X, Y)$ based on the design matrix X and the observation vector Y of β is said to be *location invariant* if

$$(1) \qquad \hat{\beta}(X, Y + Xb) = \hat{\beta}(X, Y) + b, \qquad \forall\ b \in \mathbb{R}^p.$$

It is said to be *scale invariant* if

$$(2) \qquad \hat{\beta}(X, aY) = a\hat{\beta}(X, Y), \qquad \forall\ a \in \mathbb{R},\ a \neq 0.$$

A scale estimator $s(X, Y)$ of a scale parameter γ is said to be *location invariant* if

$$(3) \qquad s(X, Y + Xb) = s(X, Y), \qquad \forall\ b \in \mathbb{R}^p.$$

It is said to be *scale invariant* if

$$(4) \qquad s(X, aY)) = |a|\ s(X,Y), \qquad \forall\ a \in \mathbb{R},\ a \neq 0.$$

Now observe that M-estimators $\hat{\Delta}$ and Δ^* of $\boldsymbol{\beta}$ of Section 4.2a are location invariant but not scale invariant. The estimators $\hat{\Delta}_1$, defined at (4.2a.13), are location and scale invariant whenever s satisfies (3) and (4). Note that if s does not satisfy (3) then $\hat{\Delta}_1$ need not be location invariant. Some of the candidates for s are

(5)
$$s := \{(n-p)^{-1} \Sigma_i (Y_i - \mathbf{x}_i' \hat{\boldsymbol{\beta}})^2\}^{1/2},$$
$$s_1 := \text{ med } \{|Y_i - \mathbf{x}_i' \hat{\boldsymbol{\beta}}|;\ 1 \le i \le n\},$$
$$s_2 := \text{ med } \{|Y_i - Y_j - (\mathbf{x}_i - \mathbf{x}_j)' \hat{\boldsymbol{\beta}}|;\ 1 \le i < j \le n\},$$

where $\hat{\boldsymbol{\beta}}$ is a preliminary estimator of $\boldsymbol{\beta}$ satisfying (1) and (2).

Estimator s^2, with $\hat{\boldsymbol{\beta}}$ as the least square estimator, is the usual estimator of the error variance, assuming it exists. It is known to be non–robust against outliers in the errors. In robustness studies one needs scale estimators that are not sensitive to outliers in the errors. Estimator s_1 has been mentioned by Huber (1981, p. 175) as one such candidate. The asymptotic properties of s_1, s_2 will be discussed shortly. Here we just mention that each of these estimators estimates a different scale parameter, but that is not a point of concern if our goal is only to have location and scale invariant M-estimators of $\boldsymbol{\beta}$.

An alternative way of having location and scale invariant M-estimators of $\boldsymbol{\beta}$ is to use simultaneous M-estimation method for estimating $\boldsymbol{\beta}$ and γ of (2.3.38) as discussed in Huber (1981). We mention here, without giving details, that it is possible to study the asymptotic joint distribution of these estimators under heteroscedastic errors by using the results of Chapter 2.

We shall now study the asymptotic distributions of s_1 and s_2 under the model (1.1.1). With F_i denoting the d.f. of e_i, $H = n^{-1} \Sigma_i F_i$, let

(6)
$$p_1(y) := H(y) - H(-y),$$

(7)
$$p_2(y) := \int [H(y+x) - H(-y+x)]\, dH(x), \qquad y \ge 0.$$

Define γ_1 and γ_2 by the relations

(8)
$$p_1(\gamma_1) = 1/2,$$

(9)
$$p_2(\gamma_2) = 1/2.$$

Note that in the case $F_i \equiv F$, γ_1 is median of the distribution of $|e_1|$ and γ_2 is median of the distribution of $|e_1 - e_2|$. In general, γ_j, p_j,

etc., depend on n, but we suppress this for the sake of convenience.

The asymptotic distribution of s_j is obtained by the usual method of connecting the event $\{s_j \le a\}$ with certain events based on certain empirical processes, as is done when studying the asymptotic distribution of the sample median, $j = 1, 2$. Accordingly, let

$$\text{(10)} \qquad S(y) := \sum_i I(|Y_i - \mathbf{x}_i'\hat{\beta}| \le y),$$
$$T(y) := \sum_{1 \le i < j \le n} \sum I(|Y_i - Y_j - (\mathbf{x}_i - \mathbf{x}_j)'\hat{\beta}| \le y), \qquad y \ge 0.$$

Then, for an $a > 0$,

$$\text{(11)} \qquad \{s_1 \le a\} = \{S(a) \ge (n+1)2^{-1}\}, \qquad n \text{ odd},$$
$$\{S(a) \ge n2^{-1}\} \subseteq \{s_1 \le a\} \subseteq \{S(a) \ge n2^{-1} - 1\}, \qquad n \text{ even}.$$

Similarly, for an $a > 0$,

$$\text{(12)} \qquad \{s_2 \le a\} = \{T(a) \ge (N+1)2^{-1}\}, \qquad N := n(n-1)/2 \text{ odd},$$
$$\{T(a) \ge N2^{-1}\} \subseteq \{s_2 \le a\} \subseteq \{T(a) \ge N2^{-1} - 1\}, \qquad N \text{ even}.$$

Thus, to study the asymptotic distributions of s_j, $j = 1, 2$, it suffices to study those of $S(y)$ and $T(y)$, $y \ge 0$.

In what follows we shall be using the notation of Chapter 2 with the following modifications. As before, we shall write S_1^o, μ_1^o etc. for S_d^o, μ_d^o etc. of (2.3.1) whenever $d_{ni} \equiv n^{-1/2}$. Moreover, in (2.3.1), we shall take

$$\text{(13)} \qquad X_i = Y_i - \mathbf{x}_i'\beta = e_i, \quad \mathbf{c}_i = \mathbf{A}\mathbf{x}_i, \qquad 1 \le i \le n.$$

With these modifications, for all $n \ge 1$,

$$\text{(14)} \qquad S(y) = S_1^o(y, \mathbf{v}) - S_1^o(-y, \mathbf{v}) = n^{-1/2} \sum_i I(|e_i - \mathbf{c}_i'\mathbf{v}| \le y),$$
$$2n^{-1}T(y) = \int [S_1^o(y+x, \mathbf{v}) - S_1^o(-y+x, \mathbf{v})]\, S_1^o(dx, \mathbf{v}) - 1, \qquad y \ge 0,$$

with probability 1, where $\mathbf{v} = \mathbf{A}^{-1}(\hat{\beta} - \beta)$. Let

$$\text{(15)} \qquad \mu_1^o(y, \mathbf{u}) = \mu_1(H(y), \mathbf{u}), \quad Y_1^o(y, \mathbf{u}) = \mathcal{Y}_1(H(y), \mathbf{u}), \qquad -\infty < y < \infty;$$
$$W(y, \mathbf{u}) = Y_1^o(y, \mathbf{u}) - Y_1^o(-y, \mathbf{u}),$$
$$K(y, \mathbf{u}) = \int [Y_1^o(y+x, \mathbf{u}) - Y_1^o(-y+x, \mathbf{u})]\, dH(x), \qquad y \ge 0, \ \mathbf{u} \in \mathbb{R}^p.$$

We shall write $W(y)$, $K(y)$ etc. for $W(y, 0)$, $K(y, 0)$ etc.

Theorem 4.3.1. *Assume that* (1.1.1) *holds with* $\mathbf{X}$ *and* $\{F_{ni}\}$ *satisfying* (NX) *and* (2.3.3). *Moreover, assume that* H *is strictly increasing for each* n *and that*

$$\text{(16)} \quad \lim_{\delta \to 0} \limsup_n \sup_{0 \le s \le 1-\delta} [H(H^{-1}(s+\delta) \pm \gamma_2) - H(H^{-1}(s) \pm \gamma_2)] = 0.$$

About $\{\hat{\beta}\}$ *assume that*

$$\text{(17)} \qquad \|A^{-1}(\hat{\beta} - \beta)\| = O_p(1).$$

Then, $\forall\ a \in \mathbb{R}$,

$$\begin{aligned}\text{(18)} \quad & P(n^{1/2}(s_1 - \gamma_1) \le a\gamma_1) \\ &= P\Big(W(\gamma_1) + n^{-1/2}\Sigma_i\, \mathbf{x}_i' A\, \{f_i(\gamma_1) - f_i(-\gamma_1)\} \cdot \mathbf{v} \\ &\qquad \ge -a \cdot \gamma_1 n^{-1} \Sigma_i\, [f_i(\gamma_1) + f_i(-\gamma_1)]\Big) + o(1),\end{aligned}$$

$$\begin{aligned}\text{(19)} \quad & P(n^{1/2}(s_2 - \gamma_2) \le a\gamma_2) \\ &= P\Big(2K(\gamma_2) + n^{-3/2} \sum_i\sum_j c_{ij} \int [f_i(\gamma_2 + x) - f_i(-\gamma_2 + x)] dF_j(x) \cdot \mathbf{v} \\ &\qquad \ge -\gamma_2 a\, n^{-1} \sum_i \int [f_i(\gamma_2 + x) + f_i(-\gamma_2 + x)]\, dH(x)\Big) + o(1).\end{aligned}$$

where $c_{ij} = (\mathbf{x}_i - \mathbf{x}_j)' A$, $1 \le i, j \le n$.

Proof. We shall give the proof of (19) only; that of (18) being similar and less involved. Fix an $a \in \mathbb{R}$ and let $Q_n(a)$ denote the left hand side of (19). Assume that n is large enough so that $a_n := (an^{-1/2} + 1)\gamma_2 > 0$. Then, by (12),

$$\begin{aligned}\text{(20)} \qquad & Q_n(a) = P(T(a_n) \ge (N+1)/2), && N \text{ odd } (N := n(n-1)/2), \\ & P(T(a_n) \ge N/2) \le Q_n(a) \le P(T(a_n) \ge N2^{-1} - 1), && N \text{ even.}\end{aligned}$$

It thus suffices to study $P(T(a_n) \ge N2^{-1} + b)$, $b \in \mathbb{R}$. Now, let

$$\begin{aligned} T_1(y) &:= n^{-1/2}(2n^{-1}T(y) + 1) - n^{1/2} p_2(y), \qquad y \ge 0, \\ k_n &:= (N + 2b)\, n^{-3/2} + n^{-1/2} - n^{1/2} p_2(a_n).\end{aligned}$$

Then, direct calculations show that

(21) $$P(T(a_n) \geq N2^{-1} + b) = P(T_1(a_n) \geq k_n).$$

We now *analyze* k_n: By (9),

$$k_n = -n^{1/2}\big(p_2(a_n) - p_2(\gamma_2)\big) + O(n^{-1/2}).$$

But

$$n^{1/2}\big(p_2(a_n) - p_2(\gamma_2)\big)$$
$$= n^{1/2}\int [\{H(a_n+x) - H(\gamma_2+x)\} - \{H(-a_n+x) - H(-\gamma_2+x)\}]\, dH(x).$$

By (2.3.3), the sequence of distributions $\{p_2\}$ is tight on $(\mathbb{R}, \mathscr{B})$, implying that $\gamma_2 = O(1)$, $n^{-1/2}\gamma_2 = o(1)$. Consequently, in view (2.3.3),

$$n^{1/2}\int \{H(\pm a_n + x) - H(\pm\gamma_2 + x)\}\, dH(x)$$
$$= a\gamma_2\, n^{-1}\Sigma_i \int f_i(\pm\gamma_2 + x)\, dH(x) + o(1),$$

and

(22) $$k_n = -a\, \gamma_2 n^{-1}\, \Sigma_i \int [f_i(\gamma_2+x) + f_i(-\gamma_2+x)]dH(x) + o(1).$$

Next, we *approximate* $T_1(a_n)$ by a sum of independent r.v.'s. The proof is similar to the one used in approximating linear rank statistics of Section 3.4. From the definition of T_1 and (14),

(23) $$T_1(y) = n^{-1/2}\int [S_1^o(y+x, \mathbf{v}) - S_1^o(-y+x, \mathbf{v})]\, S_1^o(dx, \mathbf{v}) - n^{1/2}p_2(y)$$
$$= n^{-1/2}\int [Y_1^o(y+x, \mathbf{v}) - Y_1^o(-y+x, \mathbf{v})]\, S_1^o(dx, \mathbf{v})$$
$$+ n^{-1/2}\int [\mu_1^o(y+x, \mathbf{v}) - \mu_1^o(-y+x, \mathbf{v})]\, Y_1^o(dx, \mathbf{v})$$
$$+ n^{-1/2}\int [\mu_1^o(y+x, \mathbf{v}) - \mu_1^o(-y+x, \mathbf{v})]\, \mu_1^o(dx, \mathbf{v}) - n^{1/2}p_2(y)$$
$$= E_1(y) + E_2(y) + E_3(y), \quad \text{say}.$$

But

$$\begin{aligned}(24)\qquad E_3(y) &:= n^{-1/2}\int [\mu_1^o(y+x,\,v) - \mu_1^o(-y+x,\,v)]\,\mu_1^o(dx,\,v) - n^{1/2}p_2(y)\\ &= n^{-3/2}\sum_i\sum_j \int \{F_i(y+x+c_{ij}'v) - F_i(-y+x+c_{ij}'v) - F_i(y+x)\\ &\qquad\qquad + F_i(-y+x)\}\,dF_j(x)\\ &= n^{-3/2}\sum_i\sum_j c_{ij}'\,v\int [f_i(y+x) - f_i(-y+x)]\,dF_j(x) + \bar{o}_p(1),\end{aligned}$$

by (2.3.3), (NX) and (17). In this proof, $\bar{o}_p(1)$ means $o_p(1)$ uniformly in $|y| \le k$, for every $0 < k < \infty$.

Integration by parts, (17), (2.3.25), H increasing and the fact that $\int n^{-1/2}\mu_1^o(dx,\,v) = 1$ yield that

$$\begin{aligned}(25)\qquad E_2(y) &:= n^{-1/2}\int \{\mu_1^o(y+x,\,v) - \mu_1^o(-y+x,\,v)\}\,Y_1^o(dx,\,v)\\ &= n^{-1/2}\int \{Y_1^o(y+x,\,v) - Y_1^o(-y+x,\,v)\}\,\mu_1^o(dx,\,v)\\ &= \int \{Y_1^o(y+x) - Y_1^o(-y+x)\}\,dH(x) + \bar{o}_p(1).\end{aligned}$$

Similarly,

$$(26)\qquad E_1(y) = n^{-1/2}\int \{Y_1^o(y+x) - Y_1^o(-y+x)\}\,S_1^o(dx) + \bar{o}_p(1).$$

Now observe that $n^{-1/2}S_1^o = H_n$, the ordinary empirical d.f. of the errors $\{e_i\}$. Let

$$E_{11}(y) := \int \{Y_1^o(y+x) - Y_1^o(-y+x)\}\,d\big(H_n(x) - H(x)\big) = \mathcal{Z}(y) - \mathcal{Z}(-y),$$

where

$$\mathcal{Z}(\pm y) := \int Y_1^o(\pm y+x)\,d[H_n(x) - H(x)], \qquad y \ge 0.$$

We shall show that

$$(27)\qquad \mathcal{Z}(\pm a_n) = o_p(1).$$

But

$$|\mathcal{Z}(\pm a_n) - \mathcal{Z}(\pm\gamma_2)| = |\int [\mathcal{Y}_1(H(\pm a_n+x)) - \mathcal{Y}_1(H(\pm\gamma_2+x))]d\big(H_n(x) - H(x)\big)|.$$

$$\leq 2 \sup_{|y-z|\leq|a|n^{-1/2}\gamma} |\mathcal{Y}_1(H(y)) - \mathcal{Y}_1(H(z))|.$$

$$= o_p(1), \tag{28}$$

because of (2.3.3) and Corollary 2.3.1 applied with $d_{ni} \equiv n^{-1/2}$. Thus, to prove (27), it suffices to show that

$$\mathcal{Z}(\pm\gamma_2) = o_p(1). \tag{29}$$

But

$$|\mathcal{Z}(\pm\gamma_2)| = |\int_0^1 \big[\mathcal{Y}_1(H(\pm\gamma_2+H_n^{-1}(t))) - \mathcal{Y}_1(H(\pm\gamma_2+H^{-1}(t)))\big]dt|$$

$$\leq \sup_{0\leq t\leq 1} |\mathcal{Y}_1(H(\pm\gamma_2+H^{-1}(H\,H_n^{-1}(t)))) - \mathcal{Y}_1(H(\pm\gamma_2+H^{-1}(t)))|$$

$$= o_p(1),$$

by the assumption (16), Lemma 3.4.1 and Corollary 2.3.1 applied with $d_{ni} \equiv n^{-1/2}$. This proves (27). Consequently, from (26) and an argument like (28), it follows that

$$\begin{aligned} E_1(a_n) &= \int \{Y_1^o(a_n+x) - Y_1^o(-a_n+x)\}\, dH(x) + o_p(1) \\ &= \int \{Y_1^o(\gamma_2+x) - Y_1^o(-\gamma_2+x)\}\, dH(x) + o_p(1). \end{aligned} \tag{30}$$

From (23), (24), (25), (30) and the definition (15), we obtain

$$T_1(a_n) = 2K(\gamma_2) + n^{-3/2}\sum_i\sum_j c_{ij}'A \int \{f_i(\gamma_2+x) - f_i(-\gamma_2+x)\}\, dF_j(x)\cdot v + o_p(1). \tag{31}$$

Now, from the definition of k_n and (22), it follows that the $\lim k_n$ does not depend on b. Thus the limit of the l.h.s. of (21) is the same for $b = -1, 0, 1/2$, and, in view of (21), (22) and (31), it is given by the first term on the r.h.s. of (19). □

Remark 4.3.1. Observe that, in view of (8) and (9),

$$W(\gamma_1) = n^{-1/2}\Sigma_i \{I(|e_i| \leq \gamma_1) - 1/2\},$$

$$K(\gamma_2) = \int \{H(\gamma_2+x) - H(-\gamma_2+x)\}\, dY_1^o(x)$$
$$= n^{-1/2}\Sigma_i \{H(\gamma_2 + e_i) - H(-\gamma_2 + e_i) - 1/2\}.$$

Thus, $W(\gamma_1)$ and $K(\gamma_2)$ are the sums of bounded independent centered r.v.'s and by the L-F CLT one obtains

$$\sigma_1^{-1}\, W(\gamma_1) \underset{d}{\rightarrow} N(0, 1) \text{ and } \sigma_2^{-1}\, K(\gamma_2) \underset{d}{\rightarrow} N(0, 1), \tag{32}$$

where

$$\sigma_1^2 := \text{Var } W(\gamma_1) = n^{-1}\Sigma_i \{F_i(\gamma_1) - F_i(-\gamma_1)\}\{1 - F_i(\gamma_1) + F_i(-\gamma_1)\},$$
$$\sigma_2^2 := \text{Var } K(\gamma_2) = n^{-1}\Sigma_i \int [H(\gamma_2+x) - H(-\gamma_2+x)]^2\, dF_i(x) - (1/4).$$

Remark 4.3.2. If $\{F_i\}$ are all symmetric about zero, then from (32), (18) and (19), it follows that the asymptotic distribution of s_1 and s_2 does not depend on the initial estimator $\hat{\beta}$ of β. In fact, in this case we can deduce that

$$\tau_1^{-1}\, n^{1/2}(s_1 - \gamma_1)\gamma_1^{-1} \underset{d}{\rightarrow} N(0, 1), \tag{33}$$

$$\tau_2^{-1}\, n^{1/2}(s_2 - \gamma_2)\gamma_2^{-1} \underset{d}{\rightarrow} N(0, 1),$$

where

$$\tau_1^2 := \sigma_1^2\, \{2\gamma_1\, h(\gamma_1)\}^{-2}, \quad h(x) := n^{-1}\, \Sigma_i\, f_i(x),$$
$$\tau_2^2 := \sigma_2^2\, \{\gamma_2 \int h(\gamma_2+x)\, dH(x)\}^{-2}. \qquad \Box$$

Remark 4.3.3. *i.i.d. case.* In the case $F_i \equiv F$, the asymptotic distribution of s_1 depends on $\hat{\beta}$ unless F is symmetric around zero. However, the asymptotic distribution of s_2 does not depend on $\hat{\beta}$. This is so because in this case the coefficient of $\mathbf{v}$ in (19) is

$$n^{-3/2} \sum_i \sum_j (\mathbf{x}_i - \mathbf{x}_j)'\, \mathbf{A} \int [f(\gamma_2+x) - f(-\gamma_2+x)]dF(x) = 0.$$

That the asymptotic distribution of s_2 is independent of $\hat{\beta}$ is not surprising because s_2 is essentially a symmetrized variant of s_1. We summarize this property of s_2 as

Corollary 4.3.1. *If in model* (1.1.1), $F_{ni} \equiv F$, F *satisfies* (F1), (F2) *and* X *satisfies* (NX) *then* $\tau_2^{-1} n^{1/2}(s_2 - \gamma_2) \xrightarrow[d]{} N(0, 1)$, *where*

$$\tau_2^2 = \{\int [F(\gamma_2+x)-F(-\gamma_2+x)]^2 \, dF(x) - 1/4\} \cdot \{\int f(\gamma_2+x) \, dF(x)\}^{-2}. \quad \square$$

Note that γ_2 is now the median of the distribution of $|e_1 - e_2|$. Also, observe that the condition (16) now is equivalent to

$$\sup_{0 \le s \le 1-\delta} [P(F(e_1 - y) \le s+\delta) - P(F(e_1 - y) \le s)] \longrightarrow 0 \text{ as } \delta \longrightarrow 0, \forall\, y \in \mathbb{R},$$

which is implied by the assumptions on F. $\square$

4.4. R-ESTIMATORS OF β.

Consider the model (1.1.1) and the vector of linear rank statistics

(1) $$\mathbf{T}(\mathbf{t}) := A_1 \Sigma_i (\mathbf{x}_{ni} - \bar{\mathbf{x}}_n)\varphi(R_{it}/(n+1)), \qquad \mathbf{t} \in \mathbb{R}^p,$$

where A_1 is as in (4.2a.12) and R_{it} is the rank of $Y_{ni} - \mathbf{x}_{ni}'\mathbf{t}$ among $\{Y_{nj} - \mathbf{x}_{nj}'\mathbf{t},\ 1 \le j \le n\}$.

One of the classes of R-estimators of β is defined by the relation

(2) $$\inf_{\mathbf{t}} |\mathbf{T}(\mathbf{t})|_1 = |\mathbf{T}(\hat{\beta}_1)| = \sum_{j=1}^{p} |T_j(\mathbf{t})| = 0,$$

T_j being the jth component of $\mathbf{T}$ of (1). The estimators $\hat{\beta}_1$ were initially studied by Adichie (1967) for the case $p = 1$ and by Jurečková (1971) for $p \ge 1$.

Another class of R-estimators can be defined by the relation

(3) $$\inf_{\mathbf{t}} \|\mathbf{T}(\mathbf{t})\| = \|\mathbf{T}(\hat{\beta}_2)\|.$$

Yet another class of estimators, introduced by Jaeckel (1972), are defined by the relation

(4) $$\inf_{\mathbf{t}} \mathcal{J}(\mathbf{t}) = \mathcal{J}(\hat{\beta}_3)$$

where

(5) $$\mathcal{J}(\mathbf{t}) := \Sigma_i (Y_{ni} - \mathbf{x}_{ni}'\mathbf{t})\, \varphi(R_{it}/(n+1)), \qquad \mathbf{t} \in \mathbb{R}^p.$$

Jaeckel (op. cit.) showed that for every observation vector $(Y_1, \ldots, Y_n)$ and for every $n \ge p$, $\Sigma_i\, \varphi(i/n+1) \equiv 0$ implies that $\mathcal{J}(\mathbf{t})$ is nonnegative, continuous and convex function of $\mathbf{t}$. If, in addition, X_c has the full rank p

then the set $\{\mathbf{t};\ \mathcal{J}(\mathbf{t})\le b\}$ is bounded for every $0 \le b < \infty$, where $\mathbf{X}_c$ is defined at (4.2a.11). Consequently, $\hat{\beta}_3$ exists.

Moreover, the almost everywhere derivative of $\mathcal{J}$, w.r.t. $\mathbf{t}$, is $-\mathbf{A}_1^{-1}\mathbf{T}(\mathbf{t})$. Thus, at $\hat{\beta}_3$, $\mathbf{T}$ is nearly equal to zero and hence $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\beta}_3$ are essentially the same estimators. Jaeckel showed, using the a.u.l. property of $\mathrm{T}(\mathbf{t})$ due to Jurečková (1971), that indeed $\|\mathrm{A}_1(\hat{\beta}_1 - \hat{\beta}_3)\| = o_p(1)$.

Here we shall discuss the asymptotic distribution of $\{\hat{\beta}_2\}$ under general heteroscedastic errors. The main tool is the a.u.l. Theorem 3.2.4. We shall also conclude that $\|\mathrm{A}_1(\hat{\beta}_2 - \hat{\beta}_3)\| = o_p(1)$ under (1.1.1) with general independent errors.

To begin with note that $\mathbf{T}$ of (1) is a p–vector $(\mathrm{T}_1, ..., \mathrm{T}_p)'$ where $\mathrm{T}_j(\mathbf{t})$ is a $\mathrm{T}_d(\varphi, \mathbf{u})$ – statistic of (3.1.2) with

$$
\begin{aligned}
(6)\qquad & \mathrm{X}_{ni} = \mathrm{Y}_{ni} - \mathbf{x}_{ni}'\beta,\ \ \mathbf{c}_{ni} = \mathrm{A}_1(\mathbf{x}_{ni} - \bar{\mathbf{x}}_n), \quad \mathbf{u} = \mathbf{A}_1^{-1}(\mathbf{t} - \beta),\\
& \mathrm{d}_{ni} = \mathbf{a}_{(j)}'(\mathbf{x}_{ni} - \bar{\mathbf{x}}_n),\ 1 \le i \le n;\ \ \mathbf{a}_{(j)} = \mathrm{j}^{\text{th}} \text{ column of } \mathbf{A}_1,\ \ 1 \le j \le p.
\end{aligned}
$$

Thus specializing Theorem 3.2.4 to this case readily gives

Lemma 4.4.1. *Suppose that* (1.1.1) *holds with* F_{ni} *as a d.f. of* e_{ni}, $1 \le i \le n$. *In addition, assume that*

$$
(\mathrm{NX_c})\qquad (\mathrm{X}_c{}'\mathrm{X}_c)^{-1} \textit{ exists for all } n \ge p,
$$

$$
\max_i\,(\mathbf{x}_{ni} - \bar{\mathbf{x}}_n)'(\mathrm{X}_c{}'\mathrm{X}_c)^{-1}(\mathbf{x}_{ni} - \bar{\mathbf{x}}_n) = o(1).
$$

About $\{\mathrm{F}_{ni}\}$ *assume that* H *is strictly increasing for each* n *and that* (2.2.3b), (3.2.12), (3.2.35), (3.2.36) *hold and that*

$$
(7)\qquad \lim_{\delta\to 0}\ \limsup_n \sup_{0\le s\le 1-\delta} [\mathrm{L}_j(s+\delta) - \mathrm{L}_j(s)] = 0, \qquad j = 1, ..., p
$$

where

$$
\mathrm{L}_j(s) := \Sigma_i \left(\mathbf{a}_{(j)}'(\mathbf{x}_{ni}-\bar{\mathbf{x}}_n)\right)^2 \mathrm{F}_{ni}(\mathrm{H}^{-1}(s)),\quad 0 \le s \le 1,\quad 1 \le j \le p.
$$

Then, for every $0 < \mathrm{B} < \infty$,

$$
(8)\qquad \sup_{\varphi\in\mathscr{C},\|\mathbf{A}_1^{-1}(\mathbf{t}-\beta)\|\le \mathrm{B}} \|\mathbf{T}(\mathbf{t}) - \mathbf{T}(\beta) + \mathbf{K}_n\mathrm{A}_1^{-1}(\mathbf{t}-\beta)\| = o_p(1)
$$

where

$$
\mathbf{K}_n := \mathrm{A}_1 \int_0^1 \sum_{i=1}^n (\mathbf{x}_{ni} - \tilde{\mathbf{x}}_n(s))(\mathbf{x}_{ni} - \bar{\mathbf{x}}_n)'\, q_{ni}(s)\, d\varphi(s)\, \mathrm{A}_1
$$

$$\tilde{\mathbf{x}}_n(s) := n^{-1}\sum_{i=1}^{n} \mathbf{x}_{ni}\ell_{ni}(s),$$

$\ell_{ni}(s)$ *as in* (3.2.35) *and* $q_{ni}(s) := f_{ni}(H^{-1}(s)),\ i \le i \le n,\ 0 \le s \le 1.$ □

In order to prove the asymptotic normality of $\hat{\boldsymbol{\beta}}_2$, we need to show that $\|\mathbf{A}_1^{-1}(\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta})\| = O_p(1)$. To this effect let

$$\boldsymbol{\mu} := \mathbf{A}_1 \Sigma_i (\mathbf{x}_{ni} - \bar{\mathbf{x}}_n) \int F_{ni}(H^{-1})\, d\varphi, \qquad \mathbf{S} := \mathbf{T}(\boldsymbol{\beta}) - \boldsymbol{\mu}.$$

Observe that the distribution of $(\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta})$ does not depend on $\boldsymbol{\beta}$, even when $\{e_{ni}\}$ are not identically distributed.

Lemma 4.4.2. *In addition to the assumptions of Lemma* 4.4.1 *suppose that*

$$\|\mathbf{S} + \boldsymbol{\mu}\| = O_p(1), \tag{9}$$

$$\liminf_n \inf_{\|\boldsymbol{\theta}\|=1} |\boldsymbol{\theta}'\mathbf{K}_n\boldsymbol{\theta}| \ge \alpha \ \textit{for an}\ \alpha > 0, \tag{10}$$

$$\mathbf{K}_n^{-1} \ \textit{exists for all}\ n \ge p, \quad \|\mathbf{K}_n^{-1}\| = O(1). \tag{11}$$

Then, for every $\epsilon > 0,\ 0 < z < \infty$, *there exist a* $0 < b < \infty$ *and* N_ϵ *such that*

$$P\big(\inf_{\|\mathbf{u}\|>b} \|\mathbf{T}(\mathbf{A}_1\mathbf{u} + \boldsymbol{\beta})\| \ge z\big) \ge 1 - \epsilon, \qquad n \ge N_\epsilon. \tag{12}$$

Proof. Fix an $\epsilon > 0,\ 0 < z < \infty$. Without loss of generality assume $\boldsymbol{\beta} = \mathbf{0}$. Observe that by the C–S inequality

$$\inf_{\|\mathbf{u}\|>b} \|\mathbf{T}(\mathbf{A}_1\mathbf{u})\|^2 \ge \inf_{\|\boldsymbol{\theta}\|=1,\,|r|>b} (\boldsymbol{\theta}'\mathbf{T}(r\mathbf{A}_1\boldsymbol{\theta}))^2.$$

Thus it suffices to prove that there exist a $0 < b < \infty$ and N_ϵ such that

$$P\big(\inf_{\|\boldsymbol{\theta}\|=1,\,|r|>b} (\boldsymbol{\theta}'\mathbf{T}(r\mathbf{A}_1\boldsymbol{\theta}))^2 \ge z\big) \ge 1 - \epsilon, \qquad n > N_\epsilon. \tag{13}$$

Let, for $\mathbf{t} \in \mathbb{R}^p$, $\hat{\mathbf{T}}(\mathbf{t}) := \mathbf{T}(\mathbf{0}) - \mathbf{K}_n\mathbf{A}_1^{-1}\mathbf{t}$, so that, by (8) for every $0 < B < \infty$,

$$\sup_{\|\boldsymbol{\theta}\|=1,\,|r|\le B} |\boldsymbol{\theta}'\mathbf{T}(r\mathbf{A}_1\boldsymbol{\theta}) - \boldsymbol{\theta}'\hat{\mathbf{T}}(r\mathbf{A}_1\boldsymbol{\theta})| = o_p(1). \tag{14}$$

But

$$\theta'\hat{T}(rA_1\theta) = \theta'(S+\mu) - \theta' K_n\, \theta\, r.$$

By (9), there exist a K_ϵ and an $N_{1\epsilon}$ such that

$$P(|S+\mu| \leq K_\epsilon) \geq 1-\epsilon/2, \qquad n \geq N_{1\epsilon}.$$

Choose b to satisfy

$$\text{(15)} \qquad b \geq (K_\epsilon + z^{1/2})\alpha^{-1}, \qquad \alpha \text{ as in (10).}$$

Then

$$\text{(16)} \qquad P\Big(\inf_{\|\theta\|=1,\,|r|>b} (\theta'\hat{T}(rA_1\,\theta))^2 \geq z\Big)$$

$$\geq P\Big(\|S+\mu\| \leq -z^{1/2} + b \inf_{\|\theta\|=1} |\theta' K_n\, \theta|\Big)$$

$$\geq P(\|S+\mu\| \leq K_\epsilon) \geq 1-\epsilon/2, \qquad \forall\, n \geq N_{1\epsilon}.$$

Therefore by (14) and (16) there exist N_ϵ and b as in (15) such that

$$\text{(17)} \qquad P\Big(\inf_{\|\theta\|=1,\,|r|>b} (\theta' T(rA_1\theta))^2 \geq z\Big) \geq 1-\epsilon, \qquad n \geq N_\epsilon.$$

But

$$\theta' T(rA_1\theta) = \theta' A_1 \Sigma_i (\mathbf{x}_i - \bar{\mathbf{x}})\, \varphi(R^*_{ir}/(n+1)) = \Sigma_i\, d_i\, \varphi(R^*_{ir}/(n+1)),$$

where $d_i = \theta' A_1 (\mathbf{x}_i - \bar{\mathbf{x}})$, R^*_{ir} is the rank of $Y_i - r(\mathbf{x}_i - \bar{\mathbf{x}})' A_1\, \theta$. But such a linear rank statistic is nondecreasing in r, for every θ. See, e.g., Hájek (1969; Theorem 7E, Chapter II). This together with (17) enables one to conclude (13) and hence (12). □

Theorem 4.4.1. *Suppose that* (1.1.1) *holds and that the design matrix* X *and the error d.f.'s* $\{F_{ni}\}$ *satisfy the assumptions of Lemmas* 4.4.1 *and* 4.4.2 *above. Then*

$$\text{(18)} \qquad A_1^{-1}(\hat{\beta}_2 - \beta) - K_n^{-1}\mu = K_n^{-1}S + o_p(1).$$

Proof. Follows from Lemmas 4.4.1 and 4.4.2. □

Remark 4.4.1. Arguing as in Jaeckal combined with an argument of Lemma 4.4.2, one can show that $\|A_1^{-1}(\hat{\beta}_2 - \hat{\beta}_3)\| = o_p(1)$. Consequently, under the conditions of Lemmas 4.4.1 and 4.4.2, $\hat{\beta}_2$ and the Jaeckel estimator $\hat{\beta}_3$ also satisfy (18). □

Remark 4.4.2. Consider the case when $F_{ni} \equiv F$, F a d.f. satisfying **(F1)**, **(F2)**. Then $\mu = 0$ and $S = T(\beta)$. Moreover, under (NX_c) all other assumptions of Lemmas 4.4.1 and 4.4.2 are *a priori* satisfied. Note that here

$$\ell_{ni} \equiv 1, \quad \tilde{x}_n(s) \equiv \bar{x}_n \quad \text{and} \quad K_n \equiv \int f\, d\varphi(F) \cdot I_{p\times p}.$$

Moreover, from Theorem 3.4.3 above, it follows that $S \xrightarrow[d]{} N_p(0, \sigma_\varphi^2 I_{p\times p})$, $\sigma_\varphi^2 = \int_0^1 \varphi^2(u)du - \left(\int_0^1 \varphi(u)du\right)^2$. We summarize the above discussion in

Corollary 4.4.1. *Suppose that* (1.1.1) *with* $F_{ni} \equiv F$ *holds. Suppose that* F *and* X *satisfy* **(F1)**, **(F2)**, and (NX_c). *In addition, suppose that* φ *is nondecreasing bounded on* $[0, 1]$ *and* $\int f\, d\varphi(F) > 0$. *Then*

$$A_1^{-1}(\hat{\beta}_2 - \beta) = \left\{\int f\, d\varphi(F)\right\}^{-1} T(\beta) + o_p(1). \tag{19}$$

Moreover,

$$A_1^{-1}(\hat{\beta}_2 - \beta) \xrightarrow[d]{} N(0, \tau^2 I_{p\times p}), \qquad \tau^2 = \sigma_\varphi^2 \left(\int f\, d\varphi(F)\right)^{-2}.$$ □

This result is quite general as far as the conditions on the design matrix X and F are concerned but not that general as far as the score function φ is concerned. □

Remark 4.4.2. ***Robustness against heteroscedastic gross errors.*** First , we give a working definition of qualitative robustness. Consider the model (1.1.1). Suppose that we have modeled the errors $\{e_{ni}, 1 \le i \le n\}$ to be i.i.d. F whereas their actual d.f.'s are $\{F_{ni}, 1 \le i \le n\}$. Let $P^n := \prod_{i=1}^n F$, $Q^n := \prod_{i=1}^n F_{ni}$ denote the corresponding product probability measures.

Definition 4.4.1. A sequence of estimators $\hat{\beta}$ is said to be *qualitatively robust* for β at F against Q^n if it is consistent for β under P^n and under those Q^n that satisfy $\mathcal{D}_n := \max_i \sup_y |F_{ni}(y) - F(y)| \longrightarrow 0$.

The above definition is a variant of that of Hampel (1971). One could use the notions of weak convergence on product probability spaces to give a bit more general definition. For example we could insist that the Prohorov

distance between Q^n and P^n should tend to zero instead of requiring $\mathcal{D}_n \longrightarrow 0$. We do not pursue this any further here.

The result (18) can be used to study the qualitative robustness of $\hat{\beta}_2$ against certain heteroscedastic errors. Consider, for example, the gross errors model where, for some $0 \le \delta_{ni} \le 1$, with $\max_i \delta_{ni} \longrightarrow 0$,

$$F_{ni} = (1 - \delta_{ni})\, F + \delta_{ni}\, G, \quad 1 \le i \le n,$$

and, where G is d.f. having a uniformly continuous a.e. positive density. If, in addition, $\{\delta_{ni}\}$ satisfy

$$\|A_1 \Sigma_i (\mathbf{x}_{ni} - \bar{\mathbf{x}}_n)\, \delta_{ni}\| = O(1), \tag{20}$$

then one can readily see that $\|K_n^{-1}\| = O(1)$ and $\|\mu\| = O(1)$. It follows from (18) that $\hat{\beta}_2$ is qualitatively robust against the above heteroscedastic gross errors at every F that has uniformly continuous a.e. positive density. Examples of δ_{ni} satisfying (20) would be

$$\delta_{ni} \equiv n^{-1/2} \quad \text{or} \quad \delta_{ni} = p^{-1/2}\, \|A_1 (\mathbf{x}_{ni} - \bar{\mathbf{x}}_n)\|, \quad 1 \le i \le n.$$

It may be argued that the latter choice of contaminating proportions $\{\delta_{ni}\}$ is more natural to linear regression than the former.

A similar remark is applicable to $\hat{\beta}_1$ and $\hat{\beta}_3$. □

4.5. ESTIMATION OF Q(f).

Consider the model (1.1.1) with $F_{ni} \equiv F$, where F is a d.f. with density f on $\mathbb{R}$. Define

$$Q(f) = \int f \, d\varphi(F) \tag{1}$$

where $\varphi \in \mathscr{C}$ of (3.2.1).

As is seen from Corollary 4.4.1, the parameter Q appears in the asymptotic variance of R-estimators. The complete rank analysis of the model (1.1.1) requires an estimate of Q. This estimate is used to standardize rank test statistics when carrying out the ANOVA of linear models using Jaeckal's dispersion $\mathcal{J}$ of (4.4.5). See, for example, Hettmansperger (1984) and references therein for the rank based ANOVA.

Lehmann (1963) and Sen (1966) give estimators of Q in the one and two sample location models. These estimators are given in terms of lengths of confidence intervals based on linear rank statistics. Koul (1971) extended these estimators to the multiple linear regression model (1.1.1). In this case

these estimators are given in terms of Lebesgue measures of certain confidence regions based on ranks and are hard to compute for $p > 1$.

Cheng and Serfling (1981) discuss several estimators of Q when observations are i.i.d. F, i.e., when there are no nuisance parameters. Some of these estimators are obtained by replacing f by a kernel type density estimator and F by an empirical d.f. in Q. Scheweder (1975) discusses similar estimates of Q in the one sample location model.

In this section we discuss two types of estimators of Q. Both use a kernel type density estimator of f based on the residuals and the ordinary residual empirical d.f. to estimate F. The difference is in the way the window width and the kernel are chosen. In one the window width is partially based on the data and is of the order of square root of n and the kernel is the histogram type whereas in the other the kernel and the window width are arbitrary. It will be observed that the a.u.l. result about the residual empirical process of Corollary 2.3.5 is the basic tool needed to prove the consistency of these estimators.

We begin with the class of estimators where the window width is *partly based on the data.* Define

$$p(y) := \int [F(y+x) - F(-y+x)]\, d\varphi(F(x)), \qquad y \geq 0. \tag{2}$$

Since φ is a d.f., $p(y) \equiv P(|e - e^*| \leq y)$ where e, e^* are independent r.v.'s with respective d.f.'s F and $\varphi(F)$. Consequently, under **(F1)**, the density of p at 0 is $2Q$. This suggests that an estimate of Q can be obtained by estimating the slope of p at 0.

Recall the definition of the residual empirical process $H_n(y, \mathbf{t})$ from (1.2.1). Let $\hat{\boldsymbol{\beta}}$ be an estimator of $\boldsymbol{\beta}$ and define

$$\hat{H}_n(y) := H_n(y, \hat{\boldsymbol{\beta}}), \qquad y \in \mathbb{R}. \tag{3}$$

A natural estimator of p is obtained by substituting $\hat{H}_n$ for F in p, v.i.z.,

$$\hat{p}_n(y) := \int [\hat{H}_n(y+x) - \hat{H}_n(-y+x)]\, d\varphi(\hat{H}_n(x)), \qquad y \geq 0. \tag{4}$$

Let $-\infty = \hat{e}_{(0)} < \hat{e}_{(1)} \leq \hat{e}_{(2)} \leq \dots \leq \hat{e}_{(n)} < \hat{e}_{(n+1)} = \infty$ denote the ordered residuals $\{\hat{e}_i,\ 1 \leq i \leq n\}$, where $\hat{e}_i = Y_i - \mathbf{x}_i'\hat{\boldsymbol{\beta}},\ 1 \leq i \leq n$. Since $\varphi(\hat{H}_n)$ assigns mass $\{\varphi(j/n) - \varphi((j-1)/n)\}$ to each $\hat{e}_{(j)}$ and zero mass to each of the intervals $(\hat{e}_{(j-1)}, \hat{e}_{(j)})$, $1 \leq j \leq n+1$; it readily follows that $\forall\, y \in \mathbb{R}$,

$$\hat{p}_n(y) = \sum_{j=1}^{n}\{\varphi(j/n) - \varphi((j-1)/n)\}\,[\hat{H}_n(y+\hat{e}_{(j)}) - \hat{H}_n(-y+\hat{e}_{(j)})]$$

$$(5) \qquad = n^{-1}\sum_{j=1}^{n}\{\varphi(j/n) - \varphi((j-1)/n)\}\sum_{i=1}^{n} I(|\hat{e}_{(i)} - \hat{e}_{(j)}| \le y).$$

From (5) one sees that $\hat{p}_n(y)$ has the following interpretation. For each j, one first computes the proportion of $\{\hat{e}_{(i)}\}$ falling in the interval $[-y+\hat{e}_{(j)},\ y+\hat{e}_{(j)}]$ and then $\hat{p}_n(y)$ gives the weighted average of such proportions. Formula (5) is clearly suitable for computations.

Now, if $\{h_n\}$ is a sequence of positive numbers tending to zero, an estimator of Q is given by

$$\mathcal{Q}_n = \hat{p}_n(h_n)/2h_n.$$

This estimator can be viewed from the density estimation point of view also. Consider a kernel–type density estimator f_n of f based on the residuals $\{\hat{e}_i\}$:

$$f_n(x) := (2nh_n)^{-1}\sum_{i=1}^{n} I(|x - \hat{e}_i| \le h_n),$$

which uses the window $w_n(x) = (1/2)\cdot I(|x| \le h_n)$. Then a natural estimator of Q is

$$\int f_n\, d\varphi(\hat{H}_n) = \sum_{j=1}^{n}\{\varphi(\tfrac{j}{n}) - \varphi(\tfrac{j-1}{n})\}\, f_n(\hat{e}_{(j)}) = \mathcal{Q}_n.$$

Scheweder (1975) studied the asymptotic properties of this estimator in the one sample location model. Observe that in this case the estimator of Q does not depend on the estimator of the location parameter which makes it relatively easier to derive its asymptotic properties.

In $\mathcal{Q}_n$, there is an arbitrariness due to the choice of the window width h_n. Here we recommend that h_n be determined from the spread of the data as follows. Let $0 < \alpha < 1$, t_n^α be α^{th} quantile of $\hat{p}_n$ and define the estimator $\mathcal{Q}_n^\alpha$ of Q as

$$(6) \qquad \mathcal{Q}_n^\alpha := \hat{p}_n(n^{-1/2}t_n^\alpha)/(2n^{-1/2}t_n^\alpha).$$

The quantile t_n^α is an estimator of the α^{th} quantile t^α of p. Note that if $\varphi(s) \equiv s$, then t^α is the α^{th} quantile of the distribution of $|e_1-e_2|$ and t_n^α is the α^{th} quantile of the empirical d.f. $\hat{p}_n$ of the r.v.'s $\{|\hat{e}_i-\hat{e}_j|,\ 1 \le i, j \le n\}$. Thus, e.g., $t_n^{.5} = s_2$ of (4.3.5). Similarly, if $\varphi(s) = I(s \ge 0)$

then t^{α} (t_n^{α}) is α^{th} quantile of the d.f. of $|e_1|$ (empirical d.f. of $|\hat{e}_i|$, $1 \le i \le n$). Again, here $t_n^{.5}$ would correspond to s_1 of (4.3.5). In any case, in general, t^{α} is a scale parameter in the sense of Bickel and Lehmann (1975).

The consistency of $\mathcal{Q}_n^{\alpha}$ is asserted in the following

Theorem 4.5.1. *Let* (1.1.1) *hold with* $F_{ni} \equiv F$. *In addition to* **(NX)**, **(F1)** and **(F2)**, *assume that* $\hat{\boldsymbol{\beta}}$ *is an estimator of* $\boldsymbol{\beta}$ *satisfying* (4.3.17). *Then,*

$$\sup_{\varphi \in \mathscr{C}} |\mathcal{Q}_n^{\alpha} - Q(f)| = o_p(1). \tag{7}$$

The proof of (7) will be a consequence of the following *three* lemmas.

Lemma 4.5.1. *Under the assumptions of Theorem* 4.5.1, $\forall\ 0 \le a < \infty$,

$$\sup_{\varphi \in \mathscr{C}, 0 \le z \le a} |n^{1/2}\{\hat{p}_n(n^{-1/2}z) - p(n^{-1/2}z)\}| = o_p(1). \tag{8}$$

Consequently, $\forall\ 0 \le a < \infty$,

$$\sup_{\varphi \in \mathscr{C}, 0 \le z \le a} |n^{1/2}\hat{p}_n(n^{-1/2}z) - 2zQ(f)| = o_p(1). \tag{9}$$

Proof. We shall apply Corollary 2.3.5. Let $\mathbf{v} = \mathbf{A}^{-1}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$, $\mathbf{b}_n' = n^{-1/2} \sum_i \mathbf{x}_{ni}' \mathbf{A}$. Then, from (2.3.46), (3) and (4.3.17), we obtain

$$\sup_{-\infty \le y \le \infty} |n^{1/2}\{\hat{H}_n(y) - H_n(y)\} - \mathbf{b}_n'\mathbf{v}\, f(y)| = o_p(1). \tag{10}$$

where

$$H_n(y) \equiv H_n(y, \boldsymbol{\beta}) \equiv n^{-1} \sum_{i=1}^{n} I(e_{ni} \le y), \qquad y \in \mathbb{R}.$$

Also, we will use the notation of (2.3.1) with

$$d_{ni} \equiv n^{-1/2}, \quad X_{ni} = Y_{ni} - \mathbf{x}_{ni}'\boldsymbol{\beta}, \quad F_{ni} \equiv F \text{ and } \mathbf{u} = \mathbf{0}. \tag{11}$$

Then $Y_1(t, \mathbf{0}) \equiv n^{1/2}[H_n(F^{-1}(t)) - t]$, $0 \le t \le 1$. Write $Y_1(\cdot)$ for $Y_1(\cdot, \mathbf{0})$.

Now, (10) and φ bounded imply that,

$$n^{1/2}\{\hat{p}_n(y) - p(y)\} = n^{1/2}\int \{H_n(y+x) - H_n(-y+x)\}\, d\varphi(\hat{H}_n(x))$$
$$+ b_n' \, v \int [f(y+x) - f(-y+x)]\, d\varphi(\hat{H}_n(x))$$
$$- n^{1/2}p(y) + \bar{o}_p(1)$$
$$= R_{n1}(y) + R_{n2}(y) + R_{n3}(y) + \bar{o}_p(1), \tag{12}$$

where $\bar{o}_p(1)$ stands for a sequence of random processes that converge to zero, uniformly in $-\infty \le y \le \infty$, $\varphi \in \mathscr{C}$, in probability, and where

$$R_{n1}(y) = \int \{ Y_1(F(y+x)) - Y_1(F(-y+x))\}\, d\varphi(\hat{H}_n(x))$$
$$R_{n2}(y) = b_n' \, v \int [f(y+x) - f(-y+x)]\, d\varphi(\hat{H}_n(x))$$
$$R_{n3}(y) = n^{1/2}\{\int [F(y+x) - F(-y+x)]\, d\varphi(\hat{H}_n(x))$$
$$- \int [F(y+x) - F(-y+x)]\, d\varphi(F(x))\}, \qquad y \in \mathbb{R}.$$

From (F1), (F2), the boundedness of φ, and the asymptotic continuity of Y_1, which follows from Corollary 2.2a.1, applied to the quantities given in (11), we obtain, with $k = 2a\,\|f\|_\infty$,

$$\sup_{0\le z\le a,\, \varphi\in \mathscr{C}} |R_{n1}(n^{-1/2}z)| \le \sup_{|t-s|\le kn^{-1/2}} | Y_1(t) - Y_1(s)| = o_p(1). \tag{13}$$

Again, (F1) and the boundedness of φ imply, in a routine fashion, that

$$\sup_{0\le z\le a,\, \varphi\in \mathscr{C}} |R_{n1}(n^{-1/2}z)| = o_p(1). \tag{14}$$

Now consider R_{n3}. By the MVT, (F1) and the boundedness of φ, the first term of $R_{n3}(n^{-1/2}z)$ can be written as

$$2z \int f(\xi_{xzn})\, d\varphi(\hat{H}_n(x)) = 2z \int f(x)\, d\varphi(\hat{H}_n(x)) + \bar{o}_p(1)$$

where $\{\xi_{xzn}\}$ are real numbers such that $|\xi_{xzn} - x| \le an^{-1/2}$. Do the same

with the second integral and put the two together to obtain

$$\begin{aligned} R_{n3}(n^{-1/2}z) &= 2z\{\int f \, d\varphi(\hat{H}_n) - \int f \, d\varphi(F) \} + \bar{o}_p(1) \\ &= 2z\Big\{ \int_0^1 [q(F\hat{H}_n^{-1}(t)) - q(t)] \, d\varphi(t) \Big\} + \bar{o}_p(1). \end{aligned}$$

But,

(15) $$\sup_{0 \le t \le 1} |F\hat{H}_n^{-1}(t) - t| \le n^{-1} + \sup_y |\hat{H}_n(y) - F(y)| = o_p(1)$$

by (10) and the Glivenko-Cantelli Lemma. Hence, q being uniformly continuous, we obtain

$$\sup_{0 \le z \le a,\, \varphi \in \mathscr{C}} |R_{n3}(n^{-1/2}z)| = o_p(1).$$

This together with (12) – (15) completes the proof of (8) whereas that of (9) follows from (8) and the fact that the uniform continuity of f implies that

$$\sup_{0 \le z \le a,\, \varphi \in \mathscr{C}} |n^{1/2}p(n^{-1/2}z) - 2z\, Q(f)| \longrightarrow 0. \qquad \square$$

Lemma 4.5.2. *Under the assumptions of Theorem* 4.5.1, $\forall\ y \ge 0$,

$$\sup_{\varphi \in \mathscr{C}} |\hat{p}_n(y) - p(y)| = o_p(1).$$

Proof. Proceed as in the proof of the previous lemma to rewrite

$$\hat{p}_n(y) - p(y) = \Gamma_{n1}(y) + \Gamma_{n2}(y) + \Gamma_{n3}(y) + \bar{o}_p(1)$$

where $\Gamma_{nj} = n^{-1/2}R_{nj}$, $j = 1, 2, 3$, with R_{nj} defined at (12).

By Corollary 2.2a.2 applied to the quantities given at (10), $\|Y_1\|_\infty = O_p(1)$ and hence f, φ bounded trivially imply that

$$\sup_{\varphi \in \mathscr{C},\, y \ge 0} |\Gamma_{nj}(y)| = o_p(1), \qquad j = 1, 2.$$

Now, rewrite

$$\begin{aligned} \Gamma_{n3}(y) = &\Big[\int F(y+x) \, d\varphi(\hat{H}_n(x)) - \int F(y+x) \, d\varphi(F(x))\Big] \\ &- \Big[\int F(-y+x) \, d\varphi(\hat{H}_n(x)) - \int F(-y+x) \, d\varphi(F(x))\Big] \end{aligned}$$

$$= \Gamma_n(y) + \Gamma_n(-y), \qquad \text{say.}$$

But, $\forall\ y \in \mathbb{R}$,

$$\Gamma_n(y) = \int_0^1 \{F(y + F^{-1}(F\hat{H}_n^{-1}(t))) - F(y + F^{-1}(t))\}\, d\varphi(t) = o_p(1),$$

because of (15) and because, by **(F1)** and **(F2)**, $\forall\ y \geq 0$, $F(y + F^{-1}(t))$ is uniformly continuous function of $t \in [0, 1]$. □

Lemma 4.5.3. *Under the conditions of Theorem* 4.5.1, $\forall\ \epsilon > 0$,

$$P(|t_n^\alpha - t^\alpha| \leq \epsilon t^\alpha,\ \forall\ \varphi \in \mathscr{C}) \longrightarrow 1.$$

Proof: Observe that the event $[\hat{p}_n((1-\epsilon)t^\alpha) < \alpha \leq \hat{p}_n((1+\epsilon)t^\alpha)]$ implies the event $[(1-\epsilon)t^\alpha \leq t_n^\alpha \leq (1+\epsilon)t^\alpha]$. Hence, by two applications of Lemma 4.5.2, once with $y = (1+\epsilon)t^\alpha$, and once with $y = (1-\epsilon)t^\alpha$, we obtain that

$$\liminf_n P(|t_n^\alpha - t^\alpha| \leq \epsilon\, t^\alpha,\ \forall\ \varphi \in \mathscr{C})$$
$$\geq P\big(p((1-\epsilon)t^\alpha) < \alpha \leq p((1+\epsilon)t^\alpha),\ \forall\ \varphi \in \mathscr{C}\big) = 1. \qquad □$$

Proof of Theorem 4.5.1. Clearly, $\forall\ \varphi \in \mathscr{C}$,

$$|\mathcal{Q}_n^\alpha - Q(f)| = (2t_n^\alpha)^{-1} |n^{1/2}\, \hat{p}_n(n^{-1/2}\, t_n^\alpha) - 2t_n^\alpha\, Q(f)|.$$

By Lemma 4.5.3, $\forall\ \epsilon > 0$,

$$P(0 < t_n^\alpha \leq (1 + \epsilon)t^\alpha,\ \forall\ \varphi \in \mathscr{C}) \longrightarrow 1.$$

Hence (7) follows from (9) applied with $a = (1+\epsilon)t^\alpha$, Lemma 5.4.3 and Slutsky's Theorem. □

Remark 4.5.1. The estimator $\mathcal{Q}_n^\alpha$ shifts the burden of choosing the window width to the choice of α. There does not seem to be an easy way to recommend a universal α. In an empirical study done in Koul, Sievers and McKean (1987) that investigated level and power of some rank tests in the linear regression setting, $\alpha = 0.8$ was found to be most desirable. □

Remark 4.5.2. It is an interesting theoretical exercise to see if, for some $0 < \delta < 1$, the processes $\{n^{1/2}(\mathcal{Q}_n^\alpha - Q(f)),\ \delta \leq \alpha \leq 1 - \delta\}$ converge weakly to a Gaussian process. In the case $\varphi(t) \equiv t$, Thewarapperuma (1987)

has proved, under (F1), (F2), (NX), and (4.3.17), that $\forall$ fixed $0 < \alpha < 1$, $n^{1/2}(\mathcal{Q}_n^\alpha - Q(f)) \underset{d}{\rightarrow} N(0, \sigma^2)$, where $\sigma^2 = 16\{\int f^3(x)dx - (\int f^2(x)dx)^2\}$. □

Remark 4.5.3. As mentioned earlier, $\{t_n^\alpha,\ \varphi \in \mathscr{C}\}$ provides a class of scale estimators for the class of scale parameters $\{t^\alpha,\ \varphi \in \mathscr{C}\}$. Recall that s_1 and s_2 of (4.3.5) are special cases of these estimators. The former is obtained by taking $\varphi(u) \equiv I(u \geq 0)$ and the latter by taking $\varphi(u) \equiv u$. For general interest we state a theorem below, giving asymptotic normality of these estimators. The details of proof are similar to those of Theorem 4.3.1. To state this theorem we need to introduce appropriately modified analogues of the entities defined at (4.3.15):

$$K_1(y) := \int [Y_1^o(y+x) - Y_1^o(-y+x)]\, d\varphi(F(x)),$$

$$K_2(y) := \int Y_1^o(x)\, [f(y+x) - f(-y+x)]\, \{f(x)\}^{-1}\, d\varphi(F(x)),$$

$$K(y) := K_1(y) - K_2(y), \qquad y \geq 0,$$

where Y_1^o is as (4.3.15) adapted to the i.i.d. errors setup. It is easy to check that $K(t^\alpha)$ is $n^{-1/2} \times$ {a sum of i.i.d. r.v.'s} with $E\, K(t^\alpha) \equiv 0$ and $0 < (\sigma^\alpha)^2 := \mathrm{Var}(K(t^\alpha)) < \infty$, not depending on n.

Theorem 4.5.2. *In addition to the conditions of Theorem* 4.5.1, *assume that either* $\varphi(t) = I(t \geq u)$, $0 < u < 1$, *fixed or* φ *is uniformly differentiable on* $[0, 1]$. *Then,* $\forall\ 0 < \alpha < 1$,

$$n^{1/2}(t_n^\alpha - t^\alpha) \underset{d}{\rightarrow} N(0, (\nu^\alpha)^2),$$

where

$$(\nu^\alpha)^2 := (\sigma^\alpha)^2\{t^\alpha \int [f(t^\alpha+x) + f(-t^\alpha+x)]\, d\varphi(F(x))\}^{-2}. \qquad □$$

We now turn to the ***arbitrary window width and kernel–type estimators*** of Q. Accordingly, let K be a probability density on $\mathbb{R}$, h_n be a sequence of positive numbers and $\hat{\beta}$ and $\{\hat{e}_i\}$ be as before. Define

$$\hat{f}_n(x) := (nh_n)^{-1} \sum_{i=1}^n K\big((x - \hat{e}_i)/h_n\big),$$

$$f_n(x) := (nh_n)^{-1} \sum_{i=1}^n K\big((x - e_i)/h_n\big), \qquad x \in \mathbb{R},$$

$$\hat{\mathcal{Q}}_n := \int \hat{f}_n(x)\, d\varphi(\hat{H}_n(x)).$$

Theorem 4.5.3. *Assume that the model* (1.1.1) *with* $F_{ni} \equiv F$ *holds. In addition, assume that* (**F1**), (**F2**), (**NX**) *and* (4.3.17) *and the following hold:*

(i) $h_n > 0$, $h_n \longrightarrow 0$, $n^{1/2} h_n \longrightarrow \infty$.

(ii) K *is absolutely continuous with its* a.e. *derivative* $\dot{K}$ *satisfying* $\int |\dot{K}| < \infty$.

Then,

(16) $$\sup_{\varphi \in \mathscr{C}} |\hat{\mathcal{Q}}_n - Q(f)| = o_p(1).$$

Proof. First we show $\hat{f}_n$ approximates f. This is done in several steps. To begin with, summation by parts shows that

$$\hat{f}_n(x) - f_n(x) = -h_n^{-1} \int [\hat{H}_n(x - h_n z) - H_n(x - h_n z)] \dot{K}(z)\, dz$$

so that

$$\|\hat{f}_n - f_n\|_\infty \le (n^{1/2} h_n)^{-1} \cdot \|n^{1/2}(\hat{H}_n - H_n)\|_\infty \cdot \int |\dot{K}|.$$

Hence, by (10) and the fact that $|b_n' v| = O_p(1)$ guaranteed by (4.3.17), it readily follows that

(17) $$\|\hat{f}_n - f_n\|_\infty = O_p\big((n^{1/2} h_n)^{-1}\big) = o_p(1).$$

Now, let

$$\bar{f}_n(x) := h_n^{-1} \int K\big((x - y)/h_n\big) f(y)\, dy.$$

Note that integration by parts shows that

$$\bar{f}_n(x) = -h_n^{-1} \int \dot{K}(z) F(x - h_n z)\, dz$$

so that

(18) $$\|f_n - \bar{f}_n\|_\infty \le (n^{1/2} h_n)^{-1} \cdot \|n^{1/2}[H_n - F]\|_\infty \cdot \int |\dot{K}| = o_p(1),$$

by (i) and by the fact that $\|n^{1/2}(H_n - F)\|_\infty = O_p(1)$. Moreover,

(19) $$\|\bar{f}_n - f\|_\infty \le \sup_{|y - x| \le h_n} |f(y) - f(x)| = o(1), \qquad \text{by } (\mathbf{F1}).$$

Now, consider the difference

$$\hat{Q}_n - Q(f) = \int (\hat{f}_n - f)\, d\varphi(\hat{H}_n) + \int f\, d[\varphi(\hat{H}_n) - \varphi(F)]$$

$$= D_{n1} + D_{n2}, \qquad \text{say.}$$

Let $q(t) = f(F^{-1}(t))$. Then

$$\sup_{\varphi \in \mathscr{C}} |D_{n2}| \le \sup_{0 \le t \le 1} |q(F(\hat{H}_n^{-1}(t))) - q(t)| = o_p(1)$$

by the uniform continuity of q and (15). Also,

$$\sup_{\varphi \in \mathscr{C}} |D_{n1}| \le \|\hat{f}_n - f\|_\infty = o_p(1)$$

by (17) – (19), thereby proving (16). □□

CHAPTER 5

MINIMUM DISTANCE ESTIMATORS

5.1. INTRODUCTION

The practice of obtaining estimators of parameters by minimizing a certain distance between some functions of observations and parameters has long been present in statistics. The classical examples of this method are the Least Square and the minimum Chi Square estimators.

The minimum distance estimation (m.d.e.) method, where one obtains an estimator of a parameter by minimizing some distance between the empirical d.f. and the modeled d.f., was elevated to a general method of estimation by Wolfowitz (1953, 1954, 1957). In these papers he demonstrated that compared to the maximum likelihood estimation method, the m.d.e. method yielded consistent estimators rather cheaply in several problems of varied levels of difficulty.

This methodology saw increasing research activity from the mid–seventy's when many authors demonstrated various robustness properties of certain m.d. estimators. Beran (1977) showed that in the i.i.d. setup the minimum Hellinger distance estimators, obtained by minimizing the Hellinger distance between the modeled parametric density and an empirical density estimate, are asymptotically efficient at the true model and robust against small departures from the model, where the smallness is being measured in terms of the Hellinger metric. Beran (1978) demonstrated the powerfulness of minimum Hellinger distance estimators in the one sample location model by showing that the estimators obtained by minimizing the Hellinger distance between an estimator of the density of the residual and an estimator of the density of the negative residual are qualitatively robust and adaptive for all those symmetric error distributions that have finite Fisher information.

Parr and Schucany (1979) empirically demonstrated that in certain location models several minimum distance estimators (where several comes from the type of distances chosen) are robust. Millar (1981, 1982, 1984) proved local asymptotic minimaxity of a fairly large class of m.d. estimators, using Cramer-Von Mises type distance, in the i.i.d. setup. Donoho and Liu (1988 a, b) demonstrated certain further finite sample robustness properties of a large class of m.d. estimators and certain additional advantages of using Cramer-Von Mises and Hellinger distances. All of these authors restrict their attention to the one sample setup or to the two sample location model. See Parr (1981) for additional bibliography on m.d.e. through 1980.

Little was known till the early 1980's about how to extend the above methodology to one of the most applied models, v.i.z., the multiple linear regression model (1.1.1). Given the above optimality properties in the one- and two- sample location models, it became even more desirable to extend this methodology to this model. Only after realizing that one should use the weighted, rather than the ordinary, empiricals of the residuals to define m.d. estimators was it possible to extend this methodology satisfactorily to the model (1.1.1).

The main focus of this chapter is the m.d. estimators of $\boldsymbol{\beta}$ obtained by minimizing the Cramer-Von Mises type distances involving various w.e.p.'s. Some m.d. estimators involving the supremum distance are also discussed. Most of the estimators provide appropriate extensions of their counterparts in the one- and two- sample location models.

Section 5.2 contains definitions of several m.d. estimators. Their finite sample properties and asymptotic distributions are discussed in Sections 5.3, 5.5, respectively. Section 5.4 discusses an asymptotic theory about general minimum dispersion estimators that is of broad and independent interest. It is a self contained section. Asymptotic relative efficiency and qualitative robustness of some of the m.d. estimators of Section 5.2 are discussed in Section 5.6. Some of the proposed m.d. functionals are Hellinger differentiable in the sense of Beran (1982) as is shown in Section 5.6. Consequently they are locally asymptotically minimax (l.a.m.) in the sense of Hájek – Le Cam.

5.2. DEFINITIONS OF M.D. ESTIMATORS

To motivate the following definitions of m.d. estimators of $\boldsymbol{\beta}$ of (1.1.1), first consider the one sample location model where $Y_1 - \theta, \ldots., Y_n - \theta$ are i.i.d. F, F a *known* d.f.. Let

$$F_n(y) := n^{-1}\sum_{i=1}^{n} I(Y_i \le y), \qquad y \in \mathbb{R}. \tag{1}$$

If θ is true then $EF_n(y + \theta) = F(y),\ \forall\, y\in\mathbb{R}$. This motivates one to define m.d. estimator $\hat{\theta}$ of θ by the relation

$$\hat{\theta} = \operatorname{argmin}\{T(t);\ t\in\mathbb{R}\} \tag{2}$$

where, for a $G \in \mathbb{D}\mathcal{I}(\mathbb{R})$,

$$T(t) := n\int [F_n(y + t) - F(y)]^2\, dG(y), \qquad t \in \mathbb{R}. \tag{3}$$

Observe that (2) and (3) actually define a class of estimators $\hat{\theta}$, one corresponding to each G.

Now suppose that in (1.1.1) we model the d.f. of e_{ni} to be a *known* d.f. H_{ni}, which may be different from the actual d.f. F_{ni}, $1 \le i \le n$. How should one define a m.d. estimator of $\boldsymbol{\beta}$? Any definition should reduce to $\hat{\theta}$ when (1.1.1) is reduced to the one sample location model. One possible extension is to define

$$\hat{\boldsymbol{\beta}}_1 = \operatorname{argmin}\{K_1(\mathbf{t});\ \mathbf{t} \in\mathbb{R}^p\}, \tag{4}$$

where

(5) $$K_1(\mathbf{t}) = n^{-1}\int [\sum_{i=1}^{n}\{I(Y_{ni} \le y + \mathbf{x}_{ni}'\mathbf{t}) - H_{ni}(y)\}]^2\, dG(y), \quad \mathbf{t}\in\mathbb{R}^p.$$

If in (1.1.1) we take $p = 1$, $x_{ni1} \equiv 1$ and $H_{ni} \equiv F$ then clearly it reduces to the one sample location model and $\hat{\beta}_1$ coincides with $\hat{\theta}$ of (2). But this is also true for the estimator $\hat{\beta}_{\mathbf{X}}$ defined as follows. Recall the definition of $\{V_j\}$ from (1.2.1). Define, for $y \in \mathbb{R}$, $\mathbf{t} \in \mathbb{R}^p$, $1 \le j \le p$,

(6) $$\mathcal{Z}_j(y, \mathbf{t}) := V_j(y, \mathbf{t}) - \sum_{i=1}^{n} x_{nij}\, H_{ni}(y).$$

Let

(7) $$K_{\mathbf{X}}(\mathbf{t}) := \int \boldsymbol{\mathcal{Z}}'(y, \mathbf{t})(\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{\mathcal{Z}}(y, \mathbf{t})\, dG(y), \qquad \mathbf{t} \in \mathbb{R}^p,$$

where $\boldsymbol{\mathcal{Z}}' := (\mathcal{Z}_1, \ldots., \mathcal{Z}_p)$ and define,

(8) $$\hat{\beta}_{\mathbf{X}} = \operatorname{argmin}\{K_{\mathbf{X}}(\mathbf{t}),\ \mathbf{t}\in\mathbb{R}^p\}.$$

Which of the two estimators is the right extension of $\hat{\theta}$? Since $\{V_j, 1 \le j \le p\}$ summarize the data in (1.1.1) with probability one under the continuity assumption of $\{e_{ni}, 1 \le i \le n\}$, $\hat{\beta}_{\mathbf{X}}$ should be considered the right extension of $\hat{\theta}$. In Section 5.6 we shall see that $\hat{\beta}_{\mathbf{X}}$ is asymptotically efficient among a class of estimators $\{\hat{\beta}_{\mathbf{D}}\}$ defined as follows.

Let $\mathbf{D} = ((d_{nij}))$, $1 \le i \le n$, $1 \le j \le p$, be an $n \times p$ real matrix,

(9) $$V_{jd}(y, \mathbf{t}) := \sum_{i=1}^{n} d_{nij}\, I(Y_{ni} \le y + \mathbf{x}_{ni}'\mathbf{t}), \qquad y \in \mathbb{R},\ 1 \le j \le p,$$

and

(10) $$K_{\mathbf{D}}(\mathbf{t}) := \sum_{j=1}^{p} \int [V_{jd}(y, \mathbf{t}) - \sum_{i=1}^{n} d_{nij}\, H_{ni}(y)]^2\, dG(y), \qquad \mathbf{t} \in \mathbb{R}^p.$$

Define

(11) $$\hat{\beta}_{\mathbf{D}} = \operatorname{argmin}\{K_{\mathbf{D}}(\mathbf{t}),\ \mathbf{t}\in\mathbb{R}^p\}.$$

If $\mathbf{D} = n^{-1/2}[1,0, \ldots, 0]_{n\times p}$ then $\hat{\beta}_{\mathbf{D}} = \hat{\beta}_1$ and if $\mathbf{D} = \mathbf{X}\mathbf{A}$ then $\hat{\beta}_{\mathbf{D}} = \hat{\beta}_{\mathbf{X}}$,

where $\mathbf{A}$ is as in (2.3.32). The above mentioned optimality of $\hat{\beta}_{\mathbf{X}}$ is stated and proved in Theorem 5.6a.1.

Another way to define m.d. estimators in the case the *modeled error d.f.'s are known* is as follows. Let

(12) $$M(s,y,\mathbf{t}) := n^{-1/2} \sum_{i=1}^{ns} \{I(Y_{ni} \le y) - H_{ni}(y - \mathbf{x}_{ni}'\mathbf{t})\}, \quad s \in [0, 1], \; y \in \mathbb{R},$$

(13) $$Q(\mathbf{t}) := \int_0^1 \int \{M(s,y,\mathbf{t})\}^2 \, dG(y) \, dL(s), \qquad \mathbf{t} \in \mathbb{R}^p,$$

where L is a d.f. on $[0, 1]$. Define

(14) $$\bar{\beta} = \operatorname{argmin}\{Q(\mathbf{t}), \; \mathbf{t} \in \mathbb{R}^p\}.$$

The estimator $\bar{\beta}$ with $L(s) \equiv s$ is essentially Millar's (1982) proposal.

Now suppose $\{H_{ni}\}$ are *unknown*. How should one define m.d. estimators of β in this case? Again, let us examine the one sample location model. In this case θ can not be identified unless the errors are symmetric about 0. Suppose that is the case. Then the r.v.'s $\{Y_i - \theta, \; 1 \le i \le n\}$ have the same distribution as $\{-Y_i + \theta, \; 1 \le i \le n\}$. A m.d. estimator θ^+ of θ is thus defined by the relation

(15) $$\theta^+ = \operatorname{argmin}\{T^+(t), \; t \in \mathbb{R}\}$$

where

(16) $$T^+(t) := n^{-1} \int [\sum_{i=1}^{n} \{I(Y_i \le y + t) - I(-Y_i < y - t)\}]^2 \, dG(y).$$

An extension of θ^+ to the model (1.1.1) is $\beta^+_{\mathbf{X}}$ defined by the relation

(17) $$\beta^+_{\mathbf{X}} = \operatorname{argmin}\{K^+_{\mathbf{X}}(\mathbf{t}), \; \mathbf{t} \in \mathbb{R}^p\}$$

where, for $\mathbf{t} \in \mathbb{R}^p$,

(18) $$K^+_{\mathbf{X}}(\mathbf{t}) := \int \mathbf{V}^+(y, \mathbf{t})(\mathbf{X}'\mathbf{X})^{-1}\mathbf{V}^+(y, \mathbf{t}) \, dG(y), \qquad \mathbf{V}^{+\prime} = (V_1^+, \ldots, V_p^+),$$

$$V_j^+(y, \mathbf{t}) := \sum_{i=1}^{n} x_{nij}\{I(Y_{ni} \le y + \mathbf{x}_{ni}'\mathbf{t}) - I(-Y_{ni} < y - \mathbf{x}_{ni}'\mathbf{t})\}, \; y \in \mathbb{R}, \; 1 \le j \le p.$$

More generally, a class of m.d. estimators of β can be defined as follows. Let $\mathbf{D}$ be as before. Define, for $y \in \mathbb{R}$, $1 \le j \le p$,,

(19a) $$Y_j^+(y, \mathbf{t}) := \sum_{i=1}^{n} d_{nij} \{I(Y_{ni} \le y + \mathbf{x}_{ni}'\mathbf{t}) - I(-Y_{ni} < y - \mathbf{x}_{ni}'\mathbf{t})\}.$$

Let $\mathbf{Y}_{\mathbf{D}}^{+'} = (Y_1^+, \dots, Y_p^+)$ and define

$$\text{(19b)} \qquad K_{\mathbf{D}}^+(\mathbf{t}) := \int \mathbf{Y}_{\mathbf{D}}^{+'}(y, \mathbf{t})\, \mathbf{Y}_{\mathbf{D}}^+(y, \mathbf{t})\, dG(y), \qquad \mathbf{t} \in \mathbb{R}^p.$$

and $\boldsymbol{\beta}_{\mathbf{D}}^+$ by the relation

$$\text{(20)} \qquad \boldsymbol{\beta}_{\mathbf{D}}^+ = \operatorname{argmin}\{K_{\mathbf{D}}^+(\mathbf{t}),\ \mathbf{t} \in \mathbb{R}^p\}.$$

Note that $\boldsymbol{\beta}_{\mathbf{X}}^+$ is $\boldsymbol{\beta}_{\mathbf{D}}^+$ with $\mathbf{D} = \mathbf{XA}$.

Next, suppose that the errors in (1.1.1) are modeled to be i.i.d., i.e., $H_{ni} \equiv F$ and F is *unknown* and *not necessarily symmetric.* Here, of course, the location parameter can not be estimated. However, the regression parameter vector $\boldsymbol{\beta}$ can be estimated provided the rank of $\mathbf{X}_c$ is p, where $\mathbf{X}_c$ is defined at (4.3.11). In this case a class of m.d. estimators of $\boldsymbol{\beta}$ is defined by $\hat{\boldsymbol{\beta}}_{\mathbf{D}}$ of (11) provided we assume that

$$\text{(21)} \qquad \sum_{i=1}^{n} d_{nij} = 0, \qquad 1 \le j \le p.$$

A member of this class that is of interest is $\hat{\boldsymbol{\beta}}_{\mathbf{D}}$ with $\mathbf{D} = \mathbf{X}_c\mathbf{A}_1$, $\mathbf{A}_1$ as in (4.3.11).

Another way to define m.d. estimators here is via the ranks. With R_{it} as in (3.1.1), let

$$\text{(22)} \qquad T_{jd}(s, \mathbf{t}) := \sum_{i=1}^{n} d_{nij}\, I(R_{it} \le ns), \qquad s \in [0, 1],\ 1 \le j \le p,$$

$$K_{\mathbf{D}}^*(\mathbf{t}) := \int \mathbf{T}_{\mathbf{D}}'(s, \mathbf{t})\, \mathbf{T}_{\mathbf{D}}(s, \mathbf{t})\, dL(s), \qquad \mathbf{t} \in \mathbb{R}^p,$$

where $\mathbf{T}_{\mathbf{D}}' = (T_1, \dots, T_p)$ and L is a d.f. on $[0, 1]$. Assume that $\mathbf{D}$ satisfies (21). Define

$$\text{(23)} \qquad \boldsymbol{\beta}_{\mathbf{D}}^* = \operatorname{argmin}\{K_{\mathbf{D}}^*(\mathbf{t}),\ \mathbf{t} \in \mathbb{R}^p\}.$$

Observe that $\{\hat{\boldsymbol{\beta}}_{\mathbf{D}}\}$, $\{\boldsymbol{\beta}_{\mathbf{D}}^+\}$ and $\{\bar{\boldsymbol{\beta}}\}$ are not scale invariant in the sense of (4.3.2). One way to make them so is to modify their definitions as follows. Define

$$\text{(24)} \qquad K_{\mathbf{D}}(a, \mathbf{t}) := \sum_{j=1}^{p} \int [V_{jd}(ay, \mathbf{t}) - \sum_{i=1}^{n} d_{nij}\, H_{ni}(y)]^2\, dG(y),$$

$$K_{\mathbf{D}}^+(a, \mathbf{t}) := \int \mathbf{Y}_{\mathbf{D}}^{+'}(ay, \mathbf{t})\mathbf{Y}_{\mathbf{D}}^+(ay, \mathbf{t})\, dG(y), \qquad \mathbf{t} \in \mathbb{R}^p,\ a \ge 0.$$

Now, scale invariant analogues of $\hat{\beta}_D$ and β_D^+ are defined as

$$(25)\qquad \hat{\beta}_D^o := \text{argmin}\,\{K_D(s, t),\ t\in\mathbb{R}^p\}, \quad \beta_D^{+o} := \text{argmin}\,\{K_D^+(s, t),\ t\in\mathbb{R}^p\},$$

where s is a scale estimator satisfying (4.3.3) and (4.3.4). One can modify $\{\bar{\beta}\}$ in a similar fashion to make it scale invariant. The class of estimators $\{\beta_D^*\}$ is scale invariant because the ranks are.

Now we define a m.d. estimator based on the *supremum distance* in the case the errors are correctly modeled to be i.i.d. F, F an arbitrary d.f. Here we shall *restrict* ourselves only to the *case of* $p = 1$. Define

$$(26)\qquad V_c(y, t) := \sum_{i=1}^{n}(x_i - \bar{x})\, I(Y_i \le y + tx_i), \qquad t, y \in \mathbb{R},$$

$$\mathcal{D}_n^+(t) := \sup\,\{V_c(y, t);\ y\in\mathbb{R}\},$$

$$\mathcal{D}_n^-(t) := -\inf\,\{V_c(y, t);\ y\in\mathbb{R}\},$$

$$\mathcal{D}_n(t) := \max\,\{\mathcal{D}_n^+(t), \mathcal{D}_n^-(t)\} = \sup\{|V_c(y, t)|;\ y\in\mathbb{R}\}, \quad t \in \mathbb{R}.$$

Finally, define the m.d. estimator

$$(27)\qquad \hat{\beta}_s := \text{argmin}\{\mathcal{D}_n(t);\ t\in\mathbb{R}\}.$$

Section 5.3 discusses some computational aspects including the existence and some finite sample properties of the above estimators. Section 5.5 proves the uniform asymptotic quadraticity of K_D, K_D^+, K_D^* and Q as processes in t. These results are used in Section 5.6 to study the asymptotic distributions and robustness of the above defined estimators.

5.3 FINITE SAMPLE PROPERTIES AND EXISTENCE

The purpose here is to discuss some computational aspects, the existence and the finite sample properties of the *four* classes of estimators introduced in the previous section. To facilitate this the dependence of these estimators and their defining statistics on the weight matrix D will not be exhibited in this section.

We first turn to *some computational aspects* of these estimators. To begin with, suppose that $p = 1$ and $G(y) = y$ in (5.2.10) and (5.2.11). Write $\hat{\beta}$, x_i, d_i for $\hat{\beta}$, x_{i1}, d_{i1}, respectively, $1 \le i \le n$. Then

$$(1)\quad K(t) = \int [\,\Sigma_i\, d_i\{I(Y_i \le y + x_i t) - H_i(y)\}]^2\, dy$$
$$= \Sigma_i\, \Sigma_j\, d_i d_j \int \{I(Y_i \le y+x_i t) - H_i(y)\}\{I(Y_j \le y+x_j t) - H_j(y)\}dy.$$

No further simplification of this occurs except for some special cases. One of them is the case of the one sample location model where $x_i \equiv 1$ and $H_i \equiv F$, in which case

$$K(t) = \int [\Sigma_i\, d_i\{I(Y_i \le y) - F(y-t)\}]^2\, dy.$$

Differentiating under the integral sign w.r.t. t (which can be justified under the sole assumption: F has a density f w.r.t λ) one obtains

$$\dot{K}(t) = 2\int [\Sigma_i\, d_i\{I(Y_i \le y + t) - F(y)\}\, dF(y)$$
$$= -\,2\, \Sigma_i\, d_i\{F(Y_i - t) - 1/2\}.$$

Upon taking $d_i \equiv n^{-1/2}$ one sees that in the one sample location model $\hat{\theta}$ of (5.2.2) corresponding to $G(y) = y$ is given as a solution of

$$(2)\qquad \Sigma_i\, F(Y_i - \hat{\theta}\,) = n/2.$$

Note that this $\hat{\theta}$ is precisely the m.l.e. of θ when $F(x) \equiv \{1 + \exp(-x)\}^{-1}$, i.e., when the errors have logistic distribution!

Another simplification of (1) occurs when we assume $\Sigma_i\, d_i = 0$ and $H_i \equiv F$. Fix a $t \in \mathbb{R}$ and let $c := \max\{Y_i - x_i t;\ 1 \le i \le n\}$. Then

$$(3)\qquad K(t) = \int [\Sigma_i\, d_i I(Y_i \le y + x_i t)]^2\, dy$$
$$= \Sigma_i\, \Sigma_j\, d_i d_j \int I[\max(Y_j - x_j t,\ Y_i - x_i t) \le y < c]\, dy$$
$$= -\,\Sigma_i\, \Sigma_j\, d_i d_j \max(Y_j - x_j t,\ Y_i - x_i t).$$

Using the relationship

$$(4)\qquad 2\max(a, b) = a + b + |a-b|, \qquad a, b \in \mathbb{R},$$

and the assumption $\Sigma_i\, d_i = 0$, one obtains

$$(5)\qquad K(t) = -2 \sum_{1 \le i < j \le n}\sum d_i d_j\, |\, Y_j - Y_i - (x_j - x_i)t\, |$$

If $d_i = x_i - \bar{x}$ in (5), then the corresponding $\hat{\beta}$ is asymptotically equivalent to the Wilcoxon type R-estimator of β as was shown by Williamson (1979). The result will also follow from the general asymptotic theory of Sections 5 and 6.

If $d_i = x_i - \bar{x}$, $1 \le i \le n$, and $x_i = 0$, $1 \le i \le r$; $x_i = 1$, $r+1 \le i \le n$ then (1.1.1) becomes the two sample location model and

$$K(t) = -2 \sum_{i=1}^{r} \sum_{j=r+1}^{n} |Y_j - Y_i - t| + \text{a r.v. constant in } t.$$

Consequently here $\hat{\beta} = \text{med}\{|Y_j - Y_i|,\ r+1 \le j \le n,\ 1 \le i \le r\}$, the usual Hodges–Lehmann estimator. The fact that in the two sample location model the Cramer–Von Mises type m.d. estimator of the location parameter is the Hodges–Lehmann estimator was first noted by Fine (1966).

Note that a relation like (5) is true for general p and G. That is, suppose that $p \ge 1$, $G \in \mathbb{D}\mathcal{I}(\mathbb{R})$ and (5.2.21) holds, then $\forall\ \mathbf{t} \in \mathbb{R}^p$,

$$(6) \qquad K(\mathbf{t}) = -2 \sum_{j=1}^{p} \sum\sum_{1 \le i < k \le n} d_{ij} d_{kj}\, |G((Y_k - \mathbf{x}_k'\mathbf{t})-) - G((Y_i - \mathbf{x}_i'\mathbf{t})-)|.$$

To prove this proceed as in (3) to conclude first that

$$K(\mathbf{t}) = -2 \sum_{j=1}^{p} \sum\sum_{1 \le i < k \le n} d_{ij} d_{kj}\, G(\max(Y_k - \mathbf{x}_k'\mathbf{t},\ Y_i - \mathbf{x}_i'\mathbf{t})-)$$

Now use the fact that $G((a \vee b)-) = G(a-) \vee G(b-)$, (5.2.21) and (4) to obtain (6). Clearly, formula (6) can be used to compute $\hat{\beta}$ in general.

Next *consider* K^+. To simplify the exposition, fix a $\mathbf{t} \in \mathbb{R}^p$ and let $r_i := Y_i - \mathbf{x}_i'\mathbf{t}$, $1 \le i \le n$; $b := \max\{r_i, -r_i;\ 1 \le i \le n\}$. Then from (5.2.19) we obtain

$$K^+(\mathbf{t}) = \sum_{j=1}^{p} \int [\, \Sigma_i\, d_{ij}\{I(r_i \le y) - I(-r_i < y)\}]^2\, dG(y).$$

Observe that the integrand is zero for $y > b$. Now expand the quadratic and integrate term by term, noting that G may have jumps, to obtain

$$K^+(\mathbf{t}) = \sum_{j=1}^{p} \sum_i \sum_k d_{ij} d_{kj} \Big\{ 2G(r_i \vee -r_k)\} - 2J(r_i) \\ - G((r_i \vee r_k)-) - G(-r_i \vee -r_k) \Big\},$$

where $J(y) := G(y) - G(y-)$, the jump in G at $y \in \mathbb{R}$. Once again use the fact that $G(a \vee b) = G(a) \vee G(b)$, (4), the invariance of the double sum under permutation and the definition of $\{r_i\}$ to conclude that

$$(7) \quad K^+(\mathbf{t}) = \sum_{j=1}^{p} \sum_i \sum_k d_{ij}d_{kj}\Big[|G(Y_i - \mathbf{x}_i'\mathbf{t}) - G(-Y_k + \mathbf{x}_k'\mathbf{t})| - J(Y_i - \mathbf{x}_i'\mathbf{t}) \\ -\tfrac{1}{2}\Big\{|G((Y_i - \mathbf{x}_i'\mathbf{t})-) - G((Y_k - \mathbf{x}_k'\mathbf{t})-)| \\ + |G(-Y_i + \mathbf{x}_i'\mathbf{t}) - G(-Y_k + \mathbf{x}_k'\mathbf{t})|\Big\}\Big].$$

Before proceeding further it is convenient to recall at this time the definition of symmetry for a $G \in \mathbb{D}\mathcal{I}(\mathbb{R})$.

Definition 5.3.1. An arbitrary $G \in \mathbb{D}\mathcal{I}(\mathbb{R})$, inducing a σ–finite measure on the Borel line $(\mathbb{R}, \mathscr{B})$, is said to be *symmetric* around 0 if

$$(8) \qquad |G(y) - G(x)| = |G(-x-) - G(-y-)|, \qquad \forall\ x, y \in \mathbb{R}.$$

or

$$(9) \qquad dG(y) = -\,dG(-y), \qquad \forall\ y \in \mathbb{R}.$$

If G is continuous then (8) is equivalent to

$$(10) \qquad |G(y) - G(x)| = |G(-x) - G(-y)|, \qquad \forall\ x, y \in \mathbb{R}.$$

Conversely, if (10) holds then G is symmetric around 0 and continuous.

Now suppose that G satisfies (8). Then (7) simplifies to

$$(7') \quad K^+(\mathbf{t}) = \sum_{j=1}^{p} \Sigma_i \Sigma_k\, d_{ij}d_{kj}\Big[|G(Y_i - \mathbf{x}_i'\mathbf{t}) - G(-Y_k + \mathbf{x}_k'\mathbf{t})| - J(Y_i - \mathbf{x}_i'\mathbf{t}) \\ - |G(-Y_i + \mathbf{x}_i'\mathbf{t}) - G(-Y_k + \mathbf{x}_k'\mathbf{t})|\Big].$$

And if G satisfies (10) then we obtain the relatively simpler expression

$$(7^*) \quad K^+(\mathbf{t}) = \sum_{j=1}^{p} \Sigma_i \Sigma_k\, d_{ij}d_{kj}\Big[|G(Y_i - \mathbf{x}_i'\mathbf{t}) - G(-Y_k + \mathbf{x}_k'\mathbf{t})| \\ - |G(Y_i - \mathbf{x}_i'\mathbf{t}) - G(Y_k - \mathbf{x}_k'\mathbf{t})|\Big].$$

Upon specializing (7*) to the case $G(y) = y$, $p = 1$, $d_i \equiv n^{-1/2}$ and $x_i \equiv 1$ we obtain

$$K^+(t) = n^{-1}\, \Sigma_i \Sigma_k \{|Y_i + Y_k - 2t| - |Y_i - Y_k|\}$$

and the corresponding minimizer is the well celebrated median of the pairwise means $\{(Y_i+Y_j)/2;\ 1 \le i \le j \le n\}$.

Suppose we specialize (1.1.1) to a completely randomized design with p treatments, i.e., take

$$x_{ij} = 1, \qquad m_{j-1} + 1 \le i \le m_j,$$
$$= 0, \qquad \text{otherwise},$$

where $1 \le n_j \le n$ is the j^{th} sample size, $m_0 = 0$, $m_j = n_1 + \ldots + n_j$, $1 \le j \le p$, $m_p = n$. Then, upon taking $G(y) \equiv y$, $d_{ij} \equiv x_{ij}$ in (7*), we obtain

$$K^+(\mathbf{t}) = \sum_{j=1}^{p} \sum_{i=1}^{n_j} \sum_{k=1}^{n_j} \{|Y_{ij} + Y_{kj} - 2t_j| - |Y_{ij} - Y_{kj}|\}, \qquad \mathbf{t} \in \mathbb{R}^p,$$

where Y_{ij} = the i^{th} observation from the j^{th} treatment, $1 \le j \le p$. Consequently, $\boldsymbol{\beta}^+ = (\beta_1^+, \ldots, \beta_p^+)'$, where $\beta_j^+ = \text{med}\ \{(Y_{ij} + Y_{kj})2^{-1},\ 1 \le i \le k \le n_j\}$, $1 \le j \le p$. That is, in a completely randomized design with p treatments, $\boldsymbol{\beta}^+$ corresponding to the weights $\mathbf{d}_i = \mathbf{x}_i$ and $G(y) \equiv y$ is the vector of Hodges–Lehmann estimators. Similar remark applies to the randomized block, factorial and other similar designs.

The class of estimators $\boldsymbol{\beta}^+$ also includes the well celebrated *least absolute deviation* (l.a.d.) estimator. To see this, assume that *the errors are continuous*. Choose $G = \delta_0$ – the measure degenerate at 0 – in K^+, to obtain

$$K^+(\mathbf{t}) = \sum_{j=1}^{p} \Big[\sum_{i=1}^{n} d_{ij}\ \{I(Y_i - \mathbf{x}_i'\mathbf{t} \le 0) - I(Y_i - \mathbf{x}_i'\mathbf{t} > 0)\}\Big]^2 \tag{11}$$
$$= \sum_{j=1}^{p} \Big(\sum_{i=1}^{n} d_{ij}\ \text{sgn}(Y_i - \mathbf{x}_i'\mathbf{t})\Big)^2, \text{ w.p.1}, \qquad \forall\ \mathbf{t} \in \mathbb{R}^p.$$

Upon choosing $\mathbf{d}_i \equiv \mathbf{x}_i$, one sees that the r.h.s. of (11) is precisely the square of the norm of a.e. differential of the sum of absolute deviations $\mathscr{D}(\mathbf{t}) := \Sigma_i\ |Y_i - \mathbf{x}_i'\mathbf{t}|$, $\mathbf{t} \in \mathbb{R}^p$. Clearly the minimizer of $\mathscr{D}(\mathbf{t})$ is also a minimizer of $K^+(\mathbf{t})$ of (11).

Any one of the expressions among (7), (7′) or (7*) may be used to compute $\boldsymbol{\beta}^+$ for a general G. From these expressions it becomes apparent that the computation of $\boldsymbol{\beta}^+$ is similar to the computation of maximum likelihood estimators. It is also apparent from the above discussion that both classes $\{\hat{\boldsymbol{\beta}}\}$ and $\{\boldsymbol{\beta}^+\}$ include rather interesting estimators. On the one hand we have a smooth unbounded G, v.i.z., $G(y) \equiv y$, giving rise to Hodges–Lehmann type estimators and on the other hand a highly discrete G, v.i.z., $G = \delta_0$, giving rise to the l.a.d.e.. Any large sample theory should be general enough to cover both of these cases.

We now address the question of the *existence* of these estimators in the case $p = 1$. As before when $p = 1$, we write unbold letters for scalars and d_i, x_i for d_{i1}, x_{i1}, $1 \le i \le n$. Before stating the result we need to define

$$\Gamma(y) := \Sigma_i\, I(x_i = 0)\, d_i\{I(Y_i \leq y) - I(-Y_i < y)\}, \qquad y \in \mathbb{R}.$$

Arguing as for (7) we obtain, with $b = \max\{Y_i, -Y_i;\ 1 \leq i \leq n\}$,

$$(12) \qquad \int |\Gamma|\, dG \leq \Sigma_i\, I(x_i = 0)|d_i|[G(b_-)\text{-}G(Y_{i-}) + G(b_-)\text{-}G(-Y_i)] < \infty.$$

Moreover, directly from (7) we can conclude that

$$(13) \qquad \int \Gamma^2\, dG < \infty.$$

Both (12) and (13) hold for all $n \geq 1$, for every sample $\{Y_i\}$ and for all real numbers $\{d_i\}$.

Lemma 5.3.1. *Assume that* (1.1.1) *with* $p = 1$ *holds. In addition, assume that either*

(14a) $\quad d_i x_i \geq 0, \quad \forall\ 1 \leq i \leq n,$ *or* (14b) $\quad d_i x_i \leq 0, \quad \forall\ 1 \leq i \leq n.$

Then a minimizer of K^+ *exists if either Case* 1: $G(\mathbb{R}) = \infty$, *or Case* 2: $G(\mathbb{R}) < \infty$ *and* $d_i = 0$ *whenever* $x_i = 0,\ 1 \leq i \leq n$.

If G *is continuous then a minimizer is measurable.*

Proof. The proof uses Fatou's Lemma and the D.C.T. Specialize (5.2.19) to the case $p = 1$ to obtain

$$K^+(t) = \int [\Sigma_i\, d_i\{I(Y_i \leq y + x_i t) - I(-Y_i < y - x_i t)\}]^2\, dG(y).$$

Let $\mathcal{K}^+(y, t)$ denote the integrand without the square. Then

$$\mathcal{K}^+(y, t) = \Gamma(y) + \mathcal{K}^*(y, t),$$

where

$$(15) \quad \mathcal{K}^*(y, t) = \Sigma_i\, I(x_i > 0)\, d_i\{I(Y_i \leq y + x_i t) - I(-Y_i < y - x_i t)\} + $$
$$+ \Sigma_i\, I(x_i < 0)\, d_i\{I(Y_i \leq y + x_i t) - I(-Y_i < y - x_i t)\}.$$

Clearly, $\forall\ y, t \in \mathbb{R}$,

$$|\mathcal{K}^*(y, t)| \leq \Sigma_i\, I(x_i \neq 0)|d_i| =: \alpha, \qquad \text{say.}$$

Hence

(16) $$\Gamma(y) - \alpha \le \mathcal{K}^{+}(y, t) \le \Gamma(y) + \alpha, \qquad \forall\ y, t \in \mathbb{R}.$$

Suppose that **(14a) holds**. Then, from (15) it follows that $\forall\ y \in \mathbb{R}$,

$$\mathcal{K}^{*}(y, t) \longrightarrow \pm\, \alpha \quad \text{as} \quad t \longrightarrow \pm\, \infty,$$

so that $\forall\ y \in \mathbb{R}$,

(17) $$\mathcal{K}^{+}(y, t) \longrightarrow \Gamma(y) \pm \alpha, \ \text{as}\ t \longrightarrow \pm\, \infty.$$

Now consider **Case 1**. If $\alpha = 0$ then either all $x_i \equiv 0$ or $d_i = 0$ for those i for which $x_i \neq 0$. In either case one obtains from (13) and (16) that $\forall\, t \in \mathcal{R}$, $K^{+}(t) = \int \Gamma^2\, dG < \infty$, and hence a minimizer trivially exists.

If $\alpha > 0$ then, from (12) and (13) it follows that $\int (\Gamma(y) \pm \alpha)^2\, dG(y) = \infty$, and by (16) and the Fatou Lemma, $\liminf_{t \longrightarrow \pm\infty} K^{+}(t) = \infty$. On the other hand by (7), $K^{+}(t)$ is a finite number for every real t, and hence a minimizer exists.

Next, consider **Case 2**. Here, clearly $\Gamma \equiv 0$. From (16), we obtain

$$\{\mathcal{K}^{+}(y, t)\}^2 \le \alpha^2, \qquad \forall\ y, t \in \mathbb{R},$$

and hence

$$K^{+}(t) \le \alpha^2 G(\mathbb{R}), \qquad \forall\ t \in \mathbb{R}.$$

By (17), $\mathcal{K}^{+}(y, t) \longrightarrow \pm\, \alpha$, as $t \longrightarrow \pm\, \infty$. By the D.C.T. we obtain

$$K^{+}(t) \longrightarrow \alpha^2 G(\mathbb{R}), \quad \text{as} \quad |t| \longrightarrow \infty,$$

thereby proving the existence of a minimizer of K^{+} in Case 2.

The continuity of G together with (7*) shows that K^{+} is a continuous function on $\mathbb{R}$ thereby ensuring the measurability of a minimizer, by Corollary 2.1 of Brown and Purves (1973). This completes the proof in the case of (14a). It is exactly similar when (14b) holds, hence no details will be given for that case. □

Remark 5.3.1. Observe that in some cases minimizers of K^{+} could be measurable even if G is not continuous. For example, in the case of l.a.d. estimator, G is degenerate at 0 yet a measurable minimizer exists.

The above proof is essentially due to Dhar (1991a). Dhar (1991b)

gives proofs of the existence of classes of estimators $\{\hat{\beta}\}$ and $\{\beta^+\}$ of (5.2.11) and (5.2.20) for $p \geq 1$, among other results. These proofs are somewhat complicated and will not be reproduced here. In both of these papers Dhar carries out some finite sample simulation studies and concludes that both, $\hat{\beta}$ and β^+ corresponding to $G(y) \equiv y$, show some superiority over some of the well known estimators.

Note that (14a) is *a priori* satisfied by the weights $d_i \equiv x_i$. □

Now we discuss $\bar{\beta}$ of (5.2.14). We rewrite

$$Q(\mathbf{t}) = n^{-1}\sum_i\sum_j L_{ij}\int\{I(Y_i \leq y) - H_i(y-\mathbf{x}_i'\mathbf{t})\}\{I(Y_j \leq y) - H_j(y-\mathbf{x}_j'\mathbf{t})\}dG(y)$$

where $L_{ij} = 1 - L((i\vee j)n^{-1})$, $1 \leq i, j \leq n$. Differentiating Q w.r.t. $\mathbf{t}$ under the integral sign (which can be easily justified assuming H_i has density h_i and some other mild conditions) we obtain

$$\text{(18)}\quad \dot{Q}(\mathbf{t}) = 2n^{-1}\,\Sigma_i\,\Sigma_j\,L_{ij}\int\{I(Y_i \leq y) - H_i(y - \mathbf{x}_i'\mathbf{t})\}h_i(y - \mathbf{x}_j'\mathbf{t})\,dG(y)\,\mathbf{x}_j.$$

Specialize this to the case $G(y) \equiv y$, $L(s) \equiv s$, $p = 1$, $x_i \equiv 1$ and integrate by parts, to obtain

$$\begin{aligned}\dot{Q}(t) &= -2n^{-2}\,\Sigma_i\,\Sigma_j \min(n-i, n-j)\{H_i(Y_i - t) - 1/2\}\\ &= -n^{-2}\,\Sigma_i\,(n-i)(n+i-1)\,\{H_i(Y_i - t) - 1/2\}.\end{aligned}$$

Now suppose further that $H_i \equiv F$. Then $\bar{\beta}$ is a solution t of

$$\text{(19)}\qquad \Sigma_i\,(n-i)(n+i-1)\{F(Y_i - t) - 1/2\} = 0.$$

Compare this $\bar{\beta}$ with $\hat{\theta}$ of (2). Clearly $\bar{\beta}$ given by (19) is a weighted M-estimator of the location parameter whereas $\hat{\theta}$ given by (7) is an ordinary M-estimator. Of course, if in (18) we choose $L(s) = I(s \geq 1)$, $p = 1$, $x_i \equiv 1$, $G(y) = y$ then $\hat{\theta} = \bar{\beta}$. In general $\bar{\beta}$ may be obtained as a solution of $\dot{Q}(\mathbf{t}) = \mathbf{0}$.

Next, consider β^* of (5.2.23). For the time being focus on the case $p = 1$ and $d_i \equiv x_i - \bar{x}$. Assume, without loss of generality, that the data is so arranged that $x_1 \leq x_2 \leq \ldots \leq x_n$. Let $\mathscr{S} := \{(Y_j - Y_i)/(x_j - x_i);\ i < j,\ x_i < x_j\}$, $t_0 := \min\{t;\ t\in\mathscr{S}\}$ and $t_1 := \max\{t;\ t\in\mathscr{S}\}$. Then for $x_i < x_j$, $t < t_0$ implies $t < (Y_j - Y_i)/(x_j - x_i)$ so that $R_{it} < R_{jt}$. In other words the residuals $\{Y_j - tx_j;\ 1 \leq j \leq n\}$ are naturally ordered for all $t < t_0$, w.p.1., assuming the continuity of the errors. Hence, with $T(s, t)$ denoting the $T_{1d}(s, t)$ of (22), we obtain for $t < t_0$,

$$T(s, t) = \sum_{i=1}^{k} d_i, \qquad k/n \le s < (k+1)/n, \ 1 \le k \le n-1,$$
$$= 0, \qquad 0 \le s < 1/n, \ s = 1.$$

Hence,

$$K^*(t) = \sum_{k=1}^{n-1} \omega_k \{\sum_{i=1}^{k} d_i\}^2, \qquad t < t_0.$$

where $\omega_k = [L((k+1)/n) - L(k/n)]$, $1 \le k \le n-1$. Consequently

$$K^*(t_0-) = \sum_{k=1}^{n-1} \omega_k \{\sum_{i=1}^{k} d_i\}^2.$$

Similarly using the fact $\Sigma_i d_i = 0$, one obtains

$$K^*(t) = \sum_{k=1}^{n-1} \omega_k \{\sum_{i=1}^{k} d_i\}^2 = K^*(t_{1+}), \quad t > t_1.$$

As t crosses over t_0 only one pair of adjacent residuals change their ranks. Let $x_j < x_{j+1}$ denote their respective regression constants. Then

$$K^*(t_0-) - K^*(t_0+) = \sum_{k=1}^{n-1} \omega_k \{\sum_{i=1}^{k} d_i\}^2 - $$
$$- [\sum_{\substack{k=1 \\ k \neq j}}^{n-1} \omega_k \{\sum_{i=1}^{k} d_i\}^2 + \omega_j \{d_{j+1} + \sum_{i=1}^{j-1} d_i\}^2]$$
$$= \omega_k [\{ \sum_{i=1}^{j} d_i\}^2 - \{d_{j+1} + \sum_{i=1}^{j-1} d_i\}^2].$$

But $x_1 \le x_2 \le \dots \le x_n$, $x_j < x_{j+1}$ and $\Sigma_i d_i = 0$ imply

$$\sum_{i=1}^{j} d_i < d_{k+1} + \sum_{i=1}^{j-1} d_i \le 0.$$

Hence $K^*(t_0-) > K^*(t_0+)$. Similarly it follows that $K^*(t_{1+}) > K^*(t_1-)$. Consequently, β_1 and β_2 are finite, where

$$\beta_1 := \min\{t \in \mathscr{S}, K^*(t_+) = \inf_{\Delta \in \mathscr{S}^c} K^*(\Delta)\},$$
$$\beta_2 := \max\{t \in \mathscr{S}, K^*(t_-) = \inf_{\Delta \in \mathscr{S}^c} K^*(\Delta)\},$$

and where $\mathscr{S}^c$ denotes the complement of $\mathscr{S}$. Then β^* can be uniquely defined by the relation $\beta^* = (\beta_1 + \beta_2)/2$.

This β^* corresponding to $L(s) \equiv s$ was studied by Williamson (1979, 1982). In general this estimator is asymptotically relatively more efficient than Wilcoxon type R-estimators as will be seen later on in Section 5.6.

There does not seem to be such a nice characterization for $p \geq 1$ and general $\mathbf{D}$ satisfying (5.2.21). However, proceeding as in the derivation of (6), a computational formula for K^* of (5.2.22) can be obtained to be

$$K^*(\mathbf{t}) = -2 \sum_{k=1}^{p} \Sigma_i \Sigma_j d_{ij} d_{jk} \, |L((R_{it}/n)-) - L((R_{jt}/n)-)|. \tag{20}$$

This formula is valid for a general σ–finite measure L and can be used to compute $\boldsymbol{\beta}^*$.

We now turn to the m.d. estimator defined at (5.2.26) and (5.2.27). Let $d_i \equiv x_i - \bar{x}$. The first observation one makes is that for $t \in \mathbb{R}$,

$$\mathcal{D}_n(t) := \sup_{y \in \mathbb{R}} |\sum_{i=1}^{n} d_i I(Y_i \leq y + td_i)| = \sup_{0 \leq s \leq 1} |\sum_{i=1}^{n} d_i I(R_{it} \leq ns)|.$$

Proceeding as in the above discussion pertaining to β^*, assume, without loss of generality, that the data is so arranged that $x_1 \leq x_2 \leq \dots \leq x_n$ so that $d_1 \leq d_2 \leq \dots \leq d_n$. Let $\mathscr{A}_1 := \{(Y_j - Y_i)/(d_j - d_i);\ d_i < 0,\ d_j \geq 0,\ 1 \leq i < j \leq n\}$. It can be proved that $\mathcal{D}_n^+$ ($\mathcal{D}_n^-$) is a left continuous non–decreasing (right continuous non–increasing) step function on $\mathbb{R}$ whose points of discontinuity are a subset of $\mathscr{A}_1$. Moreover, if $-\infty = t_0 < t_1 \leq t_2 \leq \dots \leq t_m < t_{m+1} = \infty$ denote the ordered members of $\mathscr{A}_1$, then $\mathcal{D}_n^+(t_1-) = 0 = \mathcal{D}_n^-(t_{m+})$ and $\mathcal{D}_n^+(t_{m+}) = \Sigma_i d_i^+ = \mathcal{D}_n^-(t_1-)$, where $d_i^+ \equiv \max(d_i, 0)$. Consequently, the following entities are finite:

$$\beta_{s1} := \inf\{t \in \mathbb{R};\ \mathcal{D}_n^+(t) \geq \mathcal{D}_n^-(t)\}, \qquad \beta_{s2} := \sup\{t \in \mathbb{R};\ \mathcal{D}_n^+(t) \leq \mathcal{D}_n^-(t)\}.$$

Note that $\beta_{s2} \geq \beta_{s1}$ w.p.1.. One can now take $\beta_s = (\beta_{s1} + \beta_{s2})/2$.

Williamson (1979) provides the proofs of the above claims and obtains the asymptotic distribution of β_s. This estimator is the precise generalization of the m.d. estimator of the two sample location parameter of Rao, Schuster and Littell (1975). Its asymptotic distribution is the same as that of their estimator.

We shall now discuss some distributional properties of the above m.d. estimators. To facilitate this discussion let $\tilde{\boldsymbol{\beta}}$ denote any one of the estimators defined at (5.2.11), (5.2.20), (5.2.23) and (5.2.27). As in Section 4.3, we shall write $\tilde{\boldsymbol{\beta}}(\mathbf{X}, \mathbf{Y})$ to emphasize the dependence on the data $\{(\mathbf{x}_i', Y_i);\ 1 \leq i \leq n\}$. It also helps to think of the defining distances K, K^+, etc. as functions of residuals. Thus we shall some times write $K(\mathbf{Y} - \mathbf{X}\mathbf{t})$ etc. for

$\mathbf{K}(\mathbf{t})$ etc. Let $\tilde{K}$ stand for either K or K^+ or K^* of (5.2.10), (5.2.19) and (5.2.22). To begin with, observe that

$$\tilde{K}(\mathbf{t}-\mathbf{b}) = \tilde{K}(\mathbf{Y}+\mathbf{X}\mathbf{b}-\mathbf{X}\mathbf{t}), \qquad \forall\ \mathbf{t}, \mathbf{b} \in \mathbb{R}^p, \tag{21}$$

so that

$$\tilde{\boldsymbol{\beta}}(\mathbf{X}, \mathbf{Y}+\mathbf{X}\mathbf{b}) = \tilde{\boldsymbol{\beta}}(\mathbf{X}, \mathbf{Y}) + \mathbf{b}, \qquad \forall\ \mathbf{b} \in \mathbb{R}^p. \tag{22}$$

Consequently, the distribution of $\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}$ does not depend on $\boldsymbol{\beta}$.

The distance measure Q of (5.2.13) does not satisfy (21) and hence the distribution of $\bar{\boldsymbol{\beta}} - \boldsymbol{\beta}$ will generally depend on $\boldsymbol{\beta}$.

In general, the classes of estimators $\{\hat{\boldsymbol{\beta}}\}$ and $\{\boldsymbol{\beta}^+\}$ are not scale invariant. However, as can be readily seen from (6) and (7), the class $\{\hat{\boldsymbol{\beta}}\}$ corresponding to $G(y) \equiv y$, $H_i \equiv F$ and those $\{\mathbf{D}\}$ that satisfy (5.2.21) and the class $\{\boldsymbol{\beta}^+\}$ corresponding to $G(y) \equiv y$ and general $\{\mathbf{D}\}$ are scale invariant in the sense of (4.3.2).

An interesting property of all of the above m.d. estimators is that they are invariant under nonsingular transformation of the design matrix $\mathbf{X}$. That is,

$$\tilde{\boldsymbol{\beta}}(\mathbf{X}\mathbf{B}, \mathbf{Y}) = \mathbf{B}^{-1}\tilde{\boldsymbol{\beta}}(\mathbf{X}, \mathbf{Y}) \text{ for every } p \times p \text{ nonsingular matrix } \mathbf{B}.$$

A similar statement holds for $\bar{\boldsymbol{\beta}}$.

We shall end this section by discussing the *symmetry* property of these estimators. In the following lemma it is implicitly assumed that all integrals involved are finite. Some sufficient conditions for that to happen will unfold as we proceed in this chapter.

Lemma 5.3.2. *Let* (1.1.1) *hold with the actual and the modeled d.f. of* e_i *equal to* H_i, $1 \le i \le n$.

(i) *If either*

(ia) $\{H_i,\ 1 \le i \le n\}$ *and* G *are symmetric around* 0 *and* $\{H_i,\ 1 \le i \le n\}$ *are continuous,*

or

(ib) $d_{ij} = -d_{n-i+1,j}$, $x_{ij} = -x_{n-i+1,j}$ *and* $H_i \equiv F$ $\forall$ $1 \le i \le n$, $1 \le j \le p$,

then

$\hat{\boldsymbol{\beta}}$ *and* $\boldsymbol{\beta}^*$ *are symmetrically distributed around* $\boldsymbol{\beta}$, *whenever they exist uniquely.*

(ii) *If* $\{H_i,\ 1 \le i \le n\}$ *and* G *are symmetric around* 0 *and either* $\{H_i,\ 1 \le i \le n\}$ *are continuous or* G *is continuous,*

then

β^+ *is symmetrically distributed around* β, *whenever it exists uniquely.*

Proof. In view of (22) there is no loss of generality in assuming that the true β is $\mathbf{0}$.

Suppose that **(ia) holds.** Then $\hat\beta(\mathbf{X}, \mathbf{Y}) \underset{d}{=} \hat\beta(\mathbf{X}, -\mathbf{Y})$. But, by definition (5.2.11), $\hat\beta(\mathbf{X}, -\mathbf{Y})$ is the minimizer of $K(-\mathbf{Y} - \mathbf{X}\mathbf{t})$ w.r.t. $\mathbf{t}$. Observe that $\forall\ \mathbf{t} \in \mathbb{R}^p$,

$$\begin{aligned} K(-\mathbf{Y} - \mathbf{X}\mathbf{t}) &= \sum_{j=1}^{p} \int [\, \Sigma_i\, d_{ij}\{I(-Y_i \le y + \mathbf{x}_i'\mathbf{t}) - H_i(y)\}]^2\, dG(y) \\ &= \sum_{j=1}^{p} \int [\, \Sigma_i\, d_{ij}\{1 - I(Y_i < -y - \mathbf{x}_i'\mathbf{t}) - H_i(y)\}]^2\, dG(y) \\ &= \sum_{j=1}^{p} \int [\, \Sigma_i\, d_{ij}\{I(Y_i < y - \mathbf{x}_i'\mathbf{t}) - H_i(y-)\}]^2\, dG(y) \end{aligned}$$

by the symmetry of $\{H_i\}$ and G. Now use the continuity of $\{H_i\}$ to conclude that, w.p.1.,

$$K(-\mathbf{Y} - \mathbf{X}\mathbf{t}) = K(\mathbf{Y} + \mathbf{X}\mathbf{t}), \qquad \forall\, \mathbf{t} \in \mathbb{R}^p,$$

so that $\hat\beta(\mathbf{X}, -\mathbf{Y}) = -\hat\beta(\mathbf{X}, \mathbf{Y})$, w.p.1, and the claim follows because $-\hat\beta(\mathbf{X}, \mathbf{Y}) = \operatorname{argmin}\{K(\mathbf{Y} + \mathbf{X}\mathbf{t});\ \mathbf{t}\in\mathbb{R}^p\}$.

Now suppose that **(ib) holds.** Then

$$\begin{aligned} &K(\mathbf{Y} + \mathbf{X}\mathbf{t}) \\ &= \sum_{j=1}^{p} \int [\Sigma_i\, d_{n-i+1,j}\{I(Y_i \le y + \mathbf{x}_{n-i+1}'\mathbf{t}) - F(y)\}]^2\, dG(y) \\ &\underset{d}{=} \sum_{j=1}^{p} \int [\Sigma_i\, d_{n-i+1,j}\{I(Y_{n-i+1} \le y + \mathbf{x}_{n-i+1}'\mathbf{t}) - F(y)\}]^2\, dG(y) \\ &= K(\mathbf{Y} - \mathbf{X}\mathbf{t}), \qquad \forall\ \mathbf{t} \in \mathbb{R}^p. \end{aligned}$$

This shows that $-\hat\beta(\mathbf{X}, \mathbf{Y}) \underset{d}{=} \hat\beta(\mathbf{X}, \mathbf{Y})$ as required. The proof for β^* is similar.

Proof of (ii). Again, $\beta^+(\mathbf{X}, \mathbf{Y}) \underset{d}{=} \beta^+(\mathbf{X}, -\mathbf{Y})$, because of the symmetry of $\{H_i\}$. But,

$$\begin{aligned}
&K^+(-\mathbf{Y} - \mathbf{X}\mathbf{t}) \\
&= \sum_{j=1}^{p} \int [\Sigma_i\, d_{ij} I(-Y_i \le y + \mathbf{x}_i'\mathbf{t}) - 1 + I(-Y_i \le -y + \mathbf{x}_i'\mathbf{t})\}]^2\, dG(y) \\
&= \sum_{j=1}^{p} \int [\Sigma_i\, d_{ij}\, \{I(Y_i < y + \mathbf{x}_i'\mathbf{t}) - 1 + I(Y_i < -y + \mathbf{x}_i'\mathbf{t})\}]^2\, dG(y) \\
&= K^+(\mathbf{Y} + \mathbf{X}\mathbf{t}), \qquad \forall\ \mathbf{t} \in \mathbb{R}^p,
\end{aligned}$$

w.p.1, if either $\{H_i\}$ or G are continuous. □

5.4. ASYMPTOTICS OF MINIMUM DISPERSION ESTIMATORS: A GENERAL CASE

This section gives a general overview of an asymptotic theory useful in inference based on minimizing an objective function of the data and parameter in general models. It is a self contained section of broad interest.

In an inferential problem consisting of a vector of n observations $\zeta_n = (\zeta_{n1}, \ldots, \zeta_{nn})'$, not necessarily independent, and a p-dimensional parameter $\boldsymbol{\theta} \in \mathbb{R}^p$, an estimator of $\boldsymbol{\theta}$ is often based on an objective function $M_n(\zeta_n, \boldsymbol{\theta})$, herein called *dispersion.* In this section an estimator of $\boldsymbol{\theta}$ obtained by minimizing $M_n(\zeta_n, \cdot)$ will be called *minimum dispersion estimator.*

Typically the sequence of dispersion M_n admits the following approximate quadratic structure. Writing $M_n(\boldsymbol{\theta})$ for $M_n(\zeta_n, \boldsymbol{\theta})$, often it turns out that $M_n(\boldsymbol{\theta}) - M_n(\boldsymbol{\theta}_0)$, under $\boldsymbol{\theta}_0$, is asymptotically like a quadratic form in $(\boldsymbol{\theta} - \boldsymbol{\theta}_0)$, for $\boldsymbol{\theta}$ close to $\boldsymbol{\theta}_0$ in a certain sense, with the coefficient of the linear term equal to a random vector which is asymptotically normally distributed. This approximation in turn is used to obtain the asymptotic distribution of minimum dispersion estimators.

The two classical examples of the above type are Gauss's least square and Fisher's maximum likelihood estimators. In the former the dispersion M_n is the error sum of squares while in the latter M_n equals $-\log L_n$, L_n denoting the likelihood function of $\boldsymbol{\theta}$ based on ζ_n. In the least squares method, $M_n(\boldsymbol{\theta}) - M_n(\boldsymbol{\theta}_0)$ is exactly quadratic in $(\boldsymbol{\theta} - \boldsymbol{\theta}_0)$, uniformly in $\boldsymbol{\theta}$ and $\boldsymbol{\theta}_0$. The random vector appearing in the linear term is typically asymptotically normally distributed. In the likelihood method, the well celebrated locally asymptotically normal (l.a.n.) models of Le Cam (1960, 1986) obey the above type of approximate quadratic structure. Other well known examples include the least absolute deviation and the minimum chi-square estimators.

The main purpose of this section is to unify the basic structure of asymptotics underlying the minimum dispersion estimators by exploiting the above type of common asymptotic quadratic structure inherent in most of the dispersions.

We now formulate general conditions for a given dispersion to be uniformly locally asymptotically quadratic (u.l.a.q.d.). Accordingly, let Ω be an open subset of $\mathbb{R}^p$ and M_n, $n \geq 1$, be a sequence of real valued functions defined on $\mathbb{R}^n \times \Omega$ such that $M_n(\cdot, \boldsymbol{\theta})$ is measurable for each $\boldsymbol{\theta}$. We shall often suppress the ζ_n coordinate in M_n and write $M_n(\boldsymbol{\theta})$ for $M_n(\zeta_n, \boldsymbol{\theta})$.

In order to state general conditions we need to define a sequence of neighborhoods $N_n(\boldsymbol{\theta}_0) := \{\boldsymbol{\theta} \in \Omega, |\boldsymbol{\delta}_n(\boldsymbol{\theta}_0)(\boldsymbol{\theta} - \boldsymbol{\theta}_0)| \leq B\}$, where $\boldsymbol{\theta}_0$ is a fixed parameter value in Ω, B is a finite number and $\{\boldsymbol{\delta}_n(\boldsymbol{\theta}_0)\}$ is a sequence of p×p symmetric positive definite matrices with norms $\|\boldsymbol{\delta}_n(\boldsymbol{\theta}_0)\|$ tending to infinity. Since $\boldsymbol{\theta}_0$ is fixed, write $\boldsymbol{\delta}_n$, N_n for $\boldsymbol{\delta}_n(\boldsymbol{\theta}_0)$, $N_n(\boldsymbol{\theta}_0)$, respectively. Similarly, let P_n denote the probability distribution of ζ_n when $\boldsymbol{\theta} = \boldsymbol{\theta}_0$.

Definition 5.4.1. A sequence of dispersions $\{M_n(\boldsymbol{\theta}),\ \boldsymbol{\theta} \in N_n\}$, $n \geq 1$, is said to be u.l.a.q. (*uniformly locally asymptotically quadratic*) if it satisfies condition $(A\tilde{1}) - (A\tilde{3})$ given below.

$(A\tilde{1})$ There exist a sequence of p×1 random vector $S_n(\boldsymbol{\theta}_0)$ and a sequence of p×p, possibly random, matrices $W_n(\boldsymbol{\theta}_0)$, such that, for every $0 < B < \infty$, and for all $\boldsymbol{\theta} \in N_n$,

$$M_n(\boldsymbol{\theta}) = M_n(\boldsymbol{\theta}_0) + (\boldsymbol{\theta} - \boldsymbol{\theta}_0)' S_n(\boldsymbol{\theta}_0) + \tfrac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)' W_n(\boldsymbol{\theta}_0)(\boldsymbol{\theta} - \boldsymbol{\theta}_0) + \bar{o}_p(1),$$

where " $\bar{o}_p(1)$" is a sequence of stochastic processes in $\boldsymbol{\theta}$ converging to zero, *uniformly in* $\boldsymbol{\theta} \in N_n$, in P_n-probability.

$(A\tilde{2})$ There exists a p×p non-singular, possibly random, matrix $\mathbf{W}(\boldsymbol{\theta}_0)$ such that

$$\boldsymbol{\delta}_n^{-1} W_n(\boldsymbol{\theta}_0) \boldsymbol{\delta}_n^{-1} = \mathbf{W}(\boldsymbol{\theta}_0) + o_p(1), \qquad (P_n).$$

$(A\tilde{3})$ There exists a p×1 r.v. $Y(\boldsymbol{\theta}_0)$ such that

$$\mathcal{L}_n\left(\boldsymbol{\delta}_n^{-1} S_n(\boldsymbol{\theta}_0),\ \boldsymbol{\delta}_n^{-1} W_n(\boldsymbol{\theta}_0) \boldsymbol{\delta}_n^{-1}\right) \Rightarrow \mathcal{L}\left(Y(\boldsymbol{\theta}_0),\ \mathbf{W}(\boldsymbol{\theta}_0)\right)$$

where $\mathcal{L}_n$, $\mathcal{L}$ denote joint probability distributions under P_n and in the limit, respectively.

Denote the conditions $(A\tilde{1})$, $(A\tilde{2})$ by (A1) and (A2), respectively, whenever $\mathbf{W}$ is *non-random* in these conditions. A sequence of dispersions $\{M_n\}$ is called *uniformly locally asymptotically* ***normal*** *quadratic* (u.l.a.n.q.) if (A1), (A2) hold and if (A3), instead of $(A\tilde{3})$, holds, where (A3) is as follows:

(A3) There exists a positive definite $p \times p$ matrix $\Sigma(\theta_0)$ such that

$$\delta_n^{-1} S_n(\theta_0) \underset{d}{\longrightarrow} N(0, \Sigma(\theta_0)), \qquad (P_n).$$

If $(\tilde{A1})$ holds *without* the *uniformity* requirement and $(\tilde{A2})$, $(\tilde{A3})$ hold then we call the given sequence M_n *locally asymptotically quadratic* (l.a.q.). If (A1) holds *without* the *uniformity* requirement and (A2), (A3) hold then the given sequence M_n is called *locally asymptotically normal quadratic.*

In the case $M_n(\theta) = -\ell n\, L_n(\theta)$, the conditions non–uniform (A1), (A2), (A3) with $\|\delta_n\| = O(n^{1/2})$, determine the well celebrated l.a.n. models of Le Cam (1960, 1986). For this particular case, $W(\theta_0)$, $\Sigma(\theta_0)$ and the limiting Fisher information matrix $F(\theta_0)$, whenever it exists, are the same.

In the above general formulation, M_n is an arbitrary dispersion satisfying $(\tilde{A1}) - (\tilde{A3})$ or (A1) – (A3). In the latter the three matrices $W(\theta_0)$, $\Sigma(\theta_0)$ and $F(\theta_0)$ are not necessarily identical. The l.a.n.q. dispersions can thus be viewed as a generalization of the l.a.n. models.

Typically in the classical i.i.d. setup the normalizing matrix δ_n is of the order square root of n whereas in the linear regression model (1.1.1) it is of the order $(X'X)^{1/2}$. In general δ_n will depend on θ_0 and is determined by the order of the asymptotic magnitude of $S_n(\theta_0)$.

An example where the full strength of $(\tilde{A1}) - (\tilde{A3})$ is realized is obtained by considering the least square dispersion in an explosive autoregression model where for some $|\rho| > 1$, $X_i = \rho X_{i-1} + e_i$, $i \geq 1$, and where $\{e_i, i \geq 1\}$ are i.i.d. r.v.'s. For details see Koul and Pflug (1990).

We now turn to the asymptotic distribution of the minimum dispersion estimators. Let $\{M_n\}$ be a sequence of u.l.a.q.d.'s. Define

$$(1) \qquad \hat{\theta}_n = \text{argmin}\{M_n(t),\ t \in \Omega\}.$$

Our goal is to investigate the asymptotic behavior of $\hat{\theta}_n$ and $M_n(\hat{\theta}_n)$. Akin to the study of the asymptotic distribution of m.l.e.'s, we must first ensure that there is a $\hat{\theta}_n$ satisfying (1) such that

$$(2) \qquad |\delta_n(\hat{\theta}_n - \theta_0)| = O_p(1).$$

Unfortunately the u.l.a.q. assumptions are not enough to guarantee (2). One set of additional assumptions that ensures (2) is the following.

(A4) $\forall\ \epsilon > 0\ \exists$ a $0 < Z_\epsilon < \infty$ and $N_{1\epsilon}$ such that

$$P_n(|M_n(\theta_0)| \leq Z_\epsilon) \geq 1 - \epsilon, \qquad \forall\ n \geq N_{1\epsilon}.$$

(A5) $\forall\ \epsilon > 0$ and $0 < \alpha < \infty$, $\exists$ an $N_{2\epsilon}$ and a b (depending on ϵ and α) such that

$$P_n\Big(\inf_{\|\delta_n(t-\theta_0)\|>b} M_n(t) \geq \alpha\Big) \geq 1-\epsilon, \qquad \forall\ n \geq N_{2\epsilon}.$$

It is convenient to let

$$Q_n(\theta, \theta_0) := (\theta-\theta_0)' S_n(\theta_0) + (1/2)(\theta-\theta_0)' W_n(\theta_0)(\theta-\theta_0), \qquad \theta \in \mathbb{R}^p,$$

and $\tilde{\theta}_n := \operatorname{argmin}\{Q_n(\theta, \theta_0),\ \theta\in\mathbb{R}^p\}$. Clearly, $\tilde{\theta}_n$ must satisfy the relation

(3) $$\mathscr{B}_n\, \delta_n(\tilde{\theta}_n-\theta_0) = -\delta_n^{-1} S_n(\theta_0).$$

where $\mathscr{B}_n := \delta_n^{-1} W_n \delta_n^{-1}$, where $W_n = W_n(\theta_0)$.

Some generality is achieved by making the following assumption.

(A6) $$\|\delta_n(\tilde{\theta}_n-\theta_0)\| = O_p(1).$$

Note that $(\mathrm{A}\tilde{2})$ and $(\mathrm{A}\tilde{3})$ imply (A6). We now state and prove

Theorem 5.4.1. *Let the dispersions* M_n *satisfy* $(\mathrm{A}\tilde{1})$, (A4) – (A6). *Then, under* P_n,

(4) $$|(\hat{\theta}_n-\tilde{\theta}_n)' \delta_n \mathscr{B}_n \delta_n(\hat{\theta}_n-\tilde{\theta}_n)| = o_p(1),$$

(5) $$\inf_{\theta\in\Omega} M_n(\theta) - M_n(\theta_0) = -(1/2)(\tilde{\theta}_n-\theta_0)' W_n(\tilde{\theta}_n-\theta_0) + o_p(1).$$

Consequently, if (A6) *is replaced by* $(\mathrm{A}\tilde{2})$ *and* $(\mathrm{A}\tilde{3})$, *then*

(6) $$\delta_n(\hat{\theta}_n-\theta_0) \xrightarrow[d]{} \{W(\theta_0)\}^{-1} Y(\theta_0),$$

and

(7) $$\inf_{\theta\in\Omega} M_n(\theta) - M_n(\theta_0) = -(1/2)\, S_n'(\theta_0)\delta_n^{-1} \mathscr{B}_n^{-1} \delta_n^{-1} S_n(\theta_0) + o_p(1).$$

If, instead of $(\mathrm{A}\tilde{1}) - (\mathrm{A}\tilde{3})$, M_n *satisfies* (A1) – (A3), *and if* (A4) *and* (A5) *hold then also* (4) – (7) *hold and*

(8) $$\delta_n(\hat{\theta}_n-\theta_0) \xrightarrow[d]{} N(0, \Gamma(\theta_0)),$$

where $\Gamma(\theta_0) = \{W(\theta_0)\}^{-1}\Sigma(\theta_0)\{W(\theta_0)\}^{-1}$.

Proof. Let Z_ϵ be as in (A4). Choose an $\alpha > Z_\epsilon$ in (A5). Then

$$\Big[|M_n(\theta_0)| \le Z_\epsilon, \inf_{|\mathbf{h}|>b} M_n(\theta_0 + \delta_n^{-1}\mathbf{h}) \ge \alpha\Big]$$

$$\subset \Big[\inf_{|\mathbf{h}|\le b} M_n(\theta_0 + \delta_n^{-1}\mathbf{h}) \le Z_\epsilon, \inf_{|\mathbf{h}|>b} M_n(\theta_0 + \delta_n^{-1}\mathbf{h}) \ge \alpha\Big]$$

$$\subset \Big[\inf_{|\mathbf{h}|>b} M_n(\theta_0 + \delta_n^{-1}\mathbf{h}) > \inf_{|\mathbf{h}|\le b} M_n(\theta_0 + \delta_n^{-1}\mathbf{h})\Big].$$

Hence by (A4) and (A5), for any $\epsilon > 0$ there exists a b (now depending only on ϵ) such that $\forall\ n \ge N_{1\epsilon} \vee N_{2\epsilon}$,

$$\text{(9)} \qquad P_n\big(\inf_{|\mathbf{h}|>b} M_n(\theta_0 + \delta_n^{-1}\mathbf{h}) > \inf_{|\mathbf{h}|\le b} M_n(\theta_0 + \delta_n^{-1}\mathbf{h})\big) \ge 1 - \epsilon,$$

This in turn ensures the validity of (2). Having verified (2), $(\tilde{A1})$ now yields

$$\text{(10)} \qquad M_n(\hat\theta_n) = M_n(\theta_0) + Q_n(\hat\theta_n, \theta_0) + o_p(1), \qquad (P_n).$$

From (A6), the inequality

$$|\inf_{\theta\in N_n} M_n(\theta) - \inf_{\theta\in N_n} [M_n(\theta_0) + Q_n(\theta, \theta_0)]|$$

$$\le \sup_{\theta\in N_n} |M_n(\theta) - [M_n(\theta_0) + Q_n(\theta, \theta_0)]|$$

and (A1), we obtain

$$\text{(11)} \qquad M_n(\hat\theta_n) = M_n(\theta_0) + Q_n(\tilde\theta_n, \theta_0) + o_p(1), \qquad (P_n).$$

Now, (10) and (11) readily yield

$$Q_n(\hat\theta_n, \theta_0) = Q_n(\tilde\theta_n, \theta_0) + o_p(1), \qquad (P_n),$$

which is precisely equivalent to the statement (4). The calim (5) follows from (3) and (11). The rest is obvious. □

Remark 5.4.1. Roughly speaking, the assumption (A5) says that the smallest value of $M_n(\theta)$ outside of N_n can be made asymptotically arbitrarily large with arbitrarily large probability. The assumption (A4) means that the sequence of r.v.'s $\{M_n(\theta_0)\}$ is bounded in probability. This assumption is usually verified by an application of the Markov inequality in the case $E_n|M_n(\theta_0)| = O(1)$, where E_n denotes the expectation under P_n. In some applications $M_n(\theta_0)$ converges weakly to a r.v. which also implies (A4). Often the verification of (A5) is rendered easy by an application of a variant of the C-S inequality. Examples of this appear in the next section when dealing with m.d. estimators of the previous section. □

We now discuss the *minimum dispersion tests of simple hypothesis*, briefly *without many details*. Consider the simple hypothesis H_0: $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. In the special case when M_n is $-\ell n\, L_n$, the likelihood ratio statistic for testing H_0 is given by $-2\inf\{M_n(\boldsymbol{\theta}) - M_n(\boldsymbol{\theta}_0);\ \boldsymbol{\theta}\in\Omega\}$. Thus, given a general dispersion function M_n, we are motivated to base a test of H_0 on the statistic

$$(12) \qquad T_n = -2\inf\{M_n(\boldsymbol{\theta}) - M_n(\boldsymbol{\theta}_0);\ \boldsymbol{\theta}\in\Omega\},$$

with large values of T_n being significant.

To study the asymptotic null distribution of T_n, note that by (7), $T_n = S_n'(\boldsymbol{\theta}_0)\boldsymbol{\delta}_n^{-1}\mathcal{B}_n^{-1}\boldsymbol{\delta}_n^{-1}S_n(\boldsymbol{\theta}_0) + o_p(1)$, (P_n). Let Y, $\mathbf{W}$ etc. stand for $Y(\boldsymbol{\theta}_0)$, $\mathbf{W}(\boldsymbol{\theta}_0)$, etc.

Proposition 5.4.1. *Under* $(\tilde{A1}) - (\tilde{A3})$, (A4), (A5), *the asymptotic null distribution of* T_n *is the same as that of* $Y'\mathbf{W}^{-1}Y$.

Under (A1) – (A5), *the asymptotic null distribution of* T_n *is the same as that of* $\mathbf{Z}'\mathbf{B}\,\mathbf{Z}$ *where* $\mathbf{Z}$ *is a* $N(0, I_{pxp})$ *r.v. and* $\mathbf{B} = \Sigma^{1/2}\mathbf{W}^{-1}\Sigma^{1/2}$. □

Remark 5.4.2. Clearly if $\mathbf{W}(\boldsymbol{\theta}_0) = \Sigma(\boldsymbol{\theta}_0)$ then the asymptotic null distribution of T_n is χ^2_p. However, if $\mathbf{W} \neq \Sigma$, the limit distribution of T_n is not a chi–square. We shall not discuss the distribution of T_n under alternatives. □

A class of examples of the u.l.a.n.q.d.'s where (A1) – (A5) are satisfied with typically $\mathbf{W} \neq \Sigma$ is given by *Huber's* M*–dispersions* for the model (1.1.1), v.i.z.,

$$M_n(\mathbf{t}) = \Sigma_i\, \rho(Y_i - \mathbf{x}_i'\mathbf{t}), \qquad \mathbf{t} \in \mathbb{R}^p,$$

where ρ is a convex function on $\mathbb{R}$ with its almost everywhere derivative ψ. As mentioned in Chapter 4 the estimators obtained by minimizing M_n are studied extensively in the literature, see Huber (1981) and references there in. These estimators include the least square and the l.a.d. estimators of $\boldsymbol{\beta}$. Now, let $g_r(t) := \int[\psi(x) - \psi(x-t)]^r dF(x)$, $t\in\mathbb{R}$, $r = 1, 2$, and suppose that F and ψ are such that $\int\psi dF = 0$, $0 < \int\psi^2 dF < \infty$, g_1 is continuously differentiable at 0 and that g_2 is continuous at 0. Then it can be shown, under (NX), that Huber's dispersion is l.a.n.q. with

$$\boldsymbol{\theta}_0 = \boldsymbol{\beta}, \quad \boldsymbol{\delta}_n = (\mathbf{X}'\mathbf{X})^{1/2}, \quad S_n(\boldsymbol{\beta}) = -\,\Sigma_i\, \mathbf{x}_i\, \psi(Y_i - \mathbf{x}_i'\boldsymbol{\beta}),$$

$$\mathbf{W}_n(\boldsymbol{\beta}) = \dot{g}(0)\mathbf{X}'\mathbf{X}, \quad \mathbf{W} = \dot{g}(0)I_{pxp}, \quad \text{and} \quad \Sigma = \int\psi^2 dF\, I_{pxp}.$$

This together with the convexity of ρ and a result in Rockafeller (1970) yields that the above dispersion is u.l.a.n.q.d. See also Heiler and Weiler (1988) and Pollard (1991).

For $\rho(x) = |x|$ and F continuous, $\psi(x) = \text{sgn}(x)$ and $g_r(t) = 2^r|F(t) - F(0)|$. The condition on g_1 now translates to the usual condition on F in terms of the density f at 0. For $\rho(x) = x^2$, $\psi(x) \equiv 2x$, $g_1(t) \equiv 2t$, so that g_1 is trivially continuously differentiable with $\dot{g}_1(0) = 2$. Note that in general $W \neq \Sigma$ unless $\dot{g}(0) = \int \psi^2\, dF$ which is the case when ψ is related to the likelihood scores.

The next section is devoted to verifying (A1) – (A5) for various dispersion introduce in Section 4.2.

5.5. ASYMPTOTIC UNIFORM QUADRATICITY

In this section we shall give sufficient conditions under which $K_{\mathbf{D}}$, $K_{\mathbf{D}}^+$, of Section 5.2 will satisfy (5.4.A1), (5.4.A4), (5.4.A5) and $K_{\mathbf{D}}^*$ and Q of Section 5.2 will satisfy (5.4.A1). As is seen from the previous section this will bring us a step closer to obtaining the asymptotic distributions of various m.d. estimators introduced in Section 5.2.

To begin with we shall focus on (5.4.A1) for $K_{\mathbf{D}}$, $K_{\mathbf{D}}^+$ and $K_{\mathbf{D}}^*$. Our basic objective is to study the asymptotic distribution of $\hat{\beta}_{\mathbf{D}}$ when the actual d.f.'s of $\{e_{ni},\ 1 \le i \le n\}$ are $\{F_{ni},\ 1 \le i \le n\}$ but we model them to be $\{H_{ni},\ 1 \le i \le n\}$. Similarly, we wish to study the asymptotic distribution of $\beta_{\mathbf{D}}^+$ when actually the errors may not be symmetric but we model them to be so. To achieve these objectives it is necessary to obtain the asymptotic results under as general a setting as possible. This of course makes the exposition that follows look somewhat complicated. The results thus obtained will enable us to study not only the asymptotic distributions of these estimators but also some of their robustness properties. With this in mind we proceed to state our assumptions.

(1) $\mathbf{X}$ satisfies (NX).

(2) With $\mathbf{d}_{(j)}$ denoting the j^{th} column of $\mathbf{D}$, $\|d_{(j)}\|^2 > 0$ for at least one j; $\|\mathbf{d}_{(j)}\|^2 = 1$ for all those j for which $\|\mathbf{d}_{(j)}\|^2 > 0$, $1 \le j \le p$.

(3) $\{F_{ni},\ 1 \le i \le n\}$ admit densities $\{f_{ni},\ 1 \le i \le n\}$ w.r.t. λ.

(4) $\{G_n\}$ is a sequence in $\mathbb{DI}(\mathbb{R})$.

(5) With $\mathbf{d}_{ni}' = (d_{ni1}, \dots, d_{nip})$, the ith row of $\mathbf{D}$, $1 \le i \le n$,

$$\int \Sigma_i \|\mathbf{d}_{ni}\|^2 F_{ni}(1-F_{ni})\, dG_n = O(1).$$

(6) With $\gamma_n := \Sigma_i \|\mathbf{d}_{ni}\|^2 f_{ni}$,

$$\limsup_n \int_{a_n}^{b_n} \int \gamma_n(y + x)\, dG_n(y)\, dx = 0$$

for any real sequences $\{a_n\}, \{b_n\}$, $a_n < b_n$, $b_n - a_n \longrightarrow 0$.

(7) With $d_{nij} = d^{+}_{nij} - d^{-}_{nij}$, $1 \le j \le p$; $\mathbf{c}_{ni} = \mathbf{A}\mathbf{x}_{ni}$, $\kappa_{ni} := \|\mathbf{c}_{ni}\|$, $1 \le i \le n$, $\forall\ \delta > 0, \forall\ \|\mathbf{v}\| \le B$,

$$\limsup_n \sum_{j=1}^{p} \int [\Sigma_i d^{\pm}_{nij}\{F_{ni}(y + \mathbf{v}'\mathbf{c}_{ni} + \delta\kappa_{ni}) - \\ - F_{ni}(y + \mathbf{v}'\mathbf{c}_{ni} - \delta\kappa_{ni})\}]^2\, dG_n(y) \le k\, \delta^2,$$

where k is a constant not depending on $\mathbf{v}$ and δ.

(8) With $\mathbf{R}_{nj} := \Sigma_i d_{nij}\, \mathbf{x}_{ni} f_{ni}$, $\nu_{nj} := \mathbf{A}\mathbf{R}_{nj}$, $1 \le j \le p$,

$$\sum_{j=1}^{p} \int \|\nu_{nj}\|^2 dG_n = O(1).$$

(9) With $\mu^{o}_{nj}(y, \mathbf{u}) := \Sigma_i d_{nij} F_{ni}(y + \mathbf{c}'_{ni}\mathbf{u})$, for each $\mathbf{u} \in \mathbb{R}^p$,

$$\sum_{j=1}^{p} \int [\mu^{o}_{nj}(y, \mathbf{u}) - \mu^{o}_{nj}(y, 0) - \mathbf{u}'\nu_{nj}(y)]^2 dG_n(y) = o(1).$$

(10) With $m_{nj} := \Sigma_i d_{nij}[F_{ni} - H_{ni}]$, $1 \le j \le p$; $\mathbf{m}'_{\mathbf{D}} = (m_{n1}, \dots, m_{np})$

$$\int \|\mathbf{m}_{\mathbf{D}}\|^2\, dG_n = O(1).$$

(11) With $\Gamma'_n(y) := (\nu_{n1}(y), \dots, \nu_{np}(y)) = \mathbf{D}'\Lambda^*(y)\mathbf{X}\mathbf{A}$, where Λ^* is defined at (4.2.1), and with $\bar{\Gamma}_n := \int \Gamma_n\, g_n\, dG_n$, where $g_n \in L_r(G_n)$, $r = 1, 2$, $n \ge 1$, is such that $g_n > 0$,

$$0 < \liminf_n \int g_n^2\, dG_n \le \limsup_n \int g_n^2\, dG_n < \infty,$$

and such that there exists an $\alpha > 0$ satisfying

$$\liminf_n \inf\{\boldsymbol{\theta}'\bar{\Gamma}_n\boldsymbol{\theta};\ \boldsymbol{\theta} \in \mathbb{R}^p, \|\boldsymbol{\theta}\| = 1\} \ge \alpha.$$

(12) Either

(a) $\boldsymbol{\theta}' d_{ni}\mathbf{x}'_{ni}\mathbf{A}\boldsymbol{\theta} \ge 0\ \ \forall\ 1 \le i \le n$ and $\forall\ \boldsymbol{\theta} \in \mathbb{R}^p, \|\boldsymbol{\theta}\| = 1$.

Or

(b) $\boldsymbol{\theta}' d_{ni}\mathbf{x}'_{ni}\mathbf{A}\boldsymbol{\theta} \le 0\ \ \forall\ 1 \le i \le n$ and $\forall\ \boldsymbol{\theta} \in \mathbb{R}^p, \|\boldsymbol{\theta}\| = 1$.

In most of the subsequent applications of the results obtained in this section, the sequence of integrating measures $\{G_n\}$ will be a fixed G. However, we formulate the results of this section in terms of sequences $\{G_n\}$ to allow extra generality. Note that if $G_n \equiv G$, $G \in \mathcal{DI}(\mathbb{R})$, then there always exists a $g \in L_r(G)$, $r = 1, 2$, such that $g > 0$, $0 < \int g^2 dG < \infty$.

Define, for $y \in \mathbb{R}$, $\mathbf{u} \in \mathbb{R}^p$, $1 \le j \le p$,

$$S_j^o(y, \mathbf{u}) = V_{jd}(y, \mathbf{A}\mathbf{u}), \qquad Y_j^o(y, \mathbf{u}) := S_j^o(y, \mathbf{u}) - \mu_j^o(y, \mathbf{u}). \tag{13}$$

Note that for each j, S_j^o, μ_j^o, Y_j^o are the same as in (2.3.2) applied to $X_{ni} = Y_{ni}$, $\mathbf{c}_{ni} = \mathbf{A}\mathbf{x}_{ni}$ and $d_{ni} = d_{nij}$, $1 \le i \le n$, $1 \le j \le p$.

Notation. For any functions $g, h : \mathbb{R}^{p+1} \longrightarrow \mathbb{R}$,

$$|g_{\mathbf{u}} - h_{\mathbf{v}}|_n^2 := \int \{g(y, \mathbf{u}) - h(y, \mathbf{v})\}^2 dG_n(y).$$

Occasionally we write $|g|_n^2$ for $|g_{\mathbf{0}}|_n^2$.

Lemma 5.5.1. *Let* $Y_{n1}, \ldots, Y_{nn}$ *be independent r.v.'s with respective d.f.'s* $F_{n1}, \ldots, F_{nn}$. *Then* (5) *implies*

$$E \sum_{j=1}^p |Y_{j\mathbf{0}}^o|_n^2 = O(1). \tag{14}$$

Proof. By Fubini's Theorem,

$$E \sum_{j=1}^p |Y_{j\mathbf{0}}|_n^2 = \int \Sigma_i \, \|\mathbf{d}_i\|^2 \, F_i(1 - F_i) \, dG_n \tag{15}$$

and hence (5) implies the Lemma. □

Lemma 5.5.2. *Let* $\{Y_{ni}\}$ *be as in Lemma* 5.5.1. *Then assumptions* (1) – (4), (6) – (10) *imply that, for every* $0 < B < \infty$,

$$E \sup_{\|\mathbf{u}\| \le B} \sum_{j=1}^p |Y_{j\mathbf{u}}^o - Y_{j\mathbf{0}}^o|_n^2 = o(1). \tag{16}$$

Proof. By Fubini's Theorem, $\forall \; \mathbf{u} \in \mathcal{N}(B)$,

$$E \sum_{j=1}^p |Y_{j\mathbf{u}}^o - Y_{j\mathbf{0}}^o|_n^2 \le \int \Sigma_i \, \|\mathbf{d}_i\|^2 |F_i(y + \mathbf{c}_i'\mathbf{u}) - F_i(y)| \, dG_n \tag{17}$$

$$\le \int_{-b_n}^{b_n} \Big(\int \gamma_n(y + x) \, dG_n(y) \Big) \, dx$$

where $b_n = B \max_i \kappa_i$, γ_n as in (6). Therefore, by assumption (6),

(18) $$E\sum_{j=1}^{p}|Y^o_{ju} - Y^o_{j0}|^2_n = o(1), \qquad \forall\ \mathbf{u}\in\mathbb{R}^p.$$

To complete the proof of (16), because of the compactness of $\mathcal{N}(B) := \{\mathbf{u}\in\mathbb{R}^p;\ \|\mathbf{u}\| \le B\}$, it suffices to show that $\forall\ \epsilon > 0\ \exists$ a $\delta > 0$ such that $\forall\ \mathbf{v}\in\mathcal{N}(B)$,

$$\limsup_n E \sup_{\|\mathbf{u}-\mathbf{v}\|\le\delta} \sum_{j=1}^{p}|L_{ju} - L_{jv}| \le \epsilon,$$

where

$$L_{ju} := |Y^o_{ju} - Y^o_{j0}|^2_n, \quad \mathbf{u}\in\mathbb{R}^p,\ 1\le j\le p.$$

Expand the quadratic, apply the C–S inequality to the cross product terms to obtain

(20) $$|L_{ju} - L_{jv}| \le |Y^o_{ju} - Y^o_{jv}|^2_n + 2|Y^o_{ju} - Y^o_{jv}|_n\,|Y^o_{jv} - Y^o_{j0}|_n,\ 1\le j\le p.$$

Moreover,

(21) $$|Y^o_{ju} - Y^o_{jv}|^2_n \le 2\{|S^o_{ju} - S^o_{jv}|^2_n + |\mu^o_{ju} - \mu^o_{jv}|^2_n\},$$

$$|S^o_{ju} - S^o_{jv}|^2_n \le 2\{|S^+_{ju} - S^+_{jv}|^2_n + |S^-_{ju} - S^-_{jv}|^2_n\},$$

$$|\mu^o_{ju} - \mu^o_{jv}|^2_n \le 2\{|\mu^+_{ju} - \mu^+_{jv}|^2_n + |\mu^-_{ju} - \mu^-_{jv}|^2_n\}, \qquad 1\le j\le p,$$

where $S^\pm_j$, $\mu^\pm_j$ are the S^o_j, μ^o_j with d_{ij} replaced by $d^\pm_{ij}$, $d^+_{ij} := \max(0, d_{ij})$, $d^-_{ij} := d^+_{ij} - d_{ij}$, $1\le i\le n$, $1\le j\le p$.

Now, $\|\mathbf{u} - \mathbf{v}\| \le \delta$, nonnegativity of $\{d^\pm_{ij}\}$, and the monotonicity of $\{F_i\}$ yields (use (2.3.15) here), that for all $1\le j\le p$,

$$|\mu^\pm_{ju} - \mu^\pm_{jv}|^2_n \le \int[\Sigma_i\, d^\pm_{ij}\{F_i(y + \mathbf{c}_i'\mathbf{v} + \delta\kappa_i) - \\ - F_i(y + \mathbf{c}_i'\mathbf{v} - \delta\kappa_i)\}]^2\, dG_n(y).$$

Therefore, by assumption (7),

(22) $$\limsup_n \sup_{\|\mathbf{u}-\mathbf{v}\|\le\delta} \sum_{j=1}^{p}|\mu^o_{ju} - \mu^o_{jv}|^2_n \le 4k\delta^2.$$

By the monotonicity of $S^\pm_j$ and (2.3.15), $\|\mathbf{u} - \mathbf{v}\| \le \delta$ implies that for all $1\le j\le p$, $y\in\mathbb{R}$,

$$- \Sigma_i d_{ij}^{\pm} I(-\delta\kappa_i < Y_i - \mathbf{v}'\mathbf{c}_i - y \le 0)$$
$$\le S_j^{\pm}(y, \mathbf{u}) - S_j^{\pm}(y, \mathbf{v})$$
$$\le \Sigma_i d_{ij}^{\pm} I(0 < Y_i - \mathbf{v}'\mathbf{c}_i - y \le \delta\kappa_i).$$

This in turn implies (using the fact that $a \le b \le c$ implies $b^2 \le a^2 + c^2$ for any reals a, b, c)

$$\begin{aligned}
&\{S_j^{\pm}(y, \mathbf{u}) - S_j^{\pm}(y, \mathbf{v})\}^2 \\
&\quad \le \{\Sigma_i d_{ij}^{\pm} I(0 < Y_i - y - \mathbf{v}'\mathbf{c}_i \le \delta\kappa_i)\}^2 + \\
&\qquad\qquad + \{\Sigma_i d_{ij}^{\pm} I(-\delta\kappa_i < Y_i - y - \mathbf{v}'\mathbf{c}_i \le 0)\}^2 \\
&\quad \le 2\,\{\Sigma_i d_{ij}^{\pm} I(-\delta\kappa_i < Y_i - y - \mathbf{v}'\mathbf{c}_i \le \delta\kappa_i)\}^2
\end{aligned}$$

for all $1 \le j \le p$ and all $y \in \mathcal{R}$. Now use the fact that for a, b real, $(a + b)^2 \le 2a^2 + 2b^2$ to conclude that, for all $1 \le j \le p$,

$$\begin{aligned}
(23)\quad &|S_{j\mathbf{u}}^{\pm} - S_{j\mathbf{v}}^{\pm}|_n^2 \\
&\le 4\int\{\Sigma_i d_{ij}^{\pm}[I(-\delta\kappa_i < Y_i - y - \mathbf{v}'\mathbf{c}_i \le \delta\kappa_i) - \\
&\qquad - p_i(y, \mathbf{v}, \delta)]\}^2\, dG_n(y) + \\
&\qquad + 4\int\{\Sigma_i d_{ij}^{\pm}\, p_i(y, \mathbf{v}, \delta)\}^2 dG_n(y) \\
&= 4\{I_j + II_j\}, \qquad \text{(say)},
\end{aligned}$$

where $p_i(y, \mathbf{v}, \delta) \equiv F_i(y + \mathbf{v}'\mathbf{c}_i + \delta\kappa_i) - F_i(y + \mathbf{v}'\mathbf{c}_i - \delta\kappa_i)$.

But $(d_{ij}^{\pm})^2 \le d_{ij}^2$ for all i and j implies that

$$\begin{aligned}
E\sum_{j=1}^{p} I_j &= \sum_{j=1}^{p}\int \Sigma_i (d_{ij}^{\pm})^2\, p_i(y, \mathbf{v}, \delta)\,(1 - p_i(y, \mathbf{v}, \delta))dG_n(y) \\
&\le \int \Sigma_i \|\mathbf{d}_i\|^2\, p_i(y, \mathbf{v}, \delta)dG_n(y) \le \int_{a_n}^{b_n}\Big(\int \gamma_n(y + s)dG_n(y)\Big)ds,
\end{aligned}$$

by (3) and Fubini, where $a_n = (-B - \delta)\max_i \kappa_i$, $b_n = (B + \delta)\max_i \kappa_i$, and where γ_n is defined in (6). Therefore, by the assumption (6),

$$(24)\qquad E\sum_{j=1}^{p} I_j = o(1).$$

From the definition of II_j in (23) and the assumption (7),

$$\limsup_n \sum_{j=1}^{p} II_j \le k\delta^2. \tag{25}$$

From (21) – (25), we obtain

$$\limsup_n E \sup_{\|\mathbf{u}-\mathbf{v}\|\le\delta} \sum_{j=1}^{p} |Y_{j\mathbf{u}}^{o} - Y_{j\mathbf{v}}^{o}|_n^2 \le 40k\,\delta^2. \tag{26}$$

Thus if we choose $0 < \delta \le (\epsilon/40k)^{1/2}$, then (19) will follow from (26), (20) and (18). This also completes the proof of (16). □

To state the next theorem we need

$$\hat{K}_{\mathbf{D}}(\mathbf{t}) := \sum_{j=1}^{p} \int \{Y_j^o(y, 0) + \mathbf{t}'\mathbf{R}_j(y) + m_j(y)\}^2 \, dG_n(y). \tag{27}$$

In (28) below, the G in $K_{\mathbf{D}}$ is assumed to have been replaced by the sequence G_n, just for extra generality.

Theorem 5.5.1. *Let* $Y_{n1}, \ldots, Y_{nn}$ *be independent r.v.'s with respective d.f.'s* $F_{n1}, \ldots, F_{nn}$. *Suppose that* $\{X, F_{ni}, H_{ni}, \mathbf{D}, G_n\}$ *satisfy* (1) – (10). *Then, for every* $0 < B < \infty$,

$$E \sup_{\|\mathbf{u}\|\le B} |K_{\mathbf{D}}(A\mathbf{u}) - \hat{K}_{\mathbf{D}}(A\mathbf{u})| = o(1). \tag{28}$$

Proof. Write $K, \hat{K}$ etc. for $K_{\mathbf{D}}, \hat{K}_{\mathbf{D}}$ etc. Note that

$$\begin{aligned} K(A\mathbf{u}) &= \sum_{j=1}^{p} \int [S_j^o(y, \mathbf{u}) - \mu_j^o(y) + m_j(y)]^2 \, dG_n(y) \\ &= \sum_{j=1}^{p} \int [Y_j^o(y, \mathbf{u}) - Y_j^o(y) + Y_j^o(y) + \mathbf{u}'\boldsymbol{\nu}_j(y) + m_j(y) \\ &\qquad + \mu_j^o(y, \mathbf{u}) - \mu_j^o(y) - \mathbf{u}'\boldsymbol{\nu}_j(y)]^2 \, dG_n(y) \end{aligned}$$

where $Y_j^o(y) \equiv Y_j^o(y, 0)$, $\mu_j^o(y) \equiv \mu_j^o(y, 0)$. Expand the quadratic and use the C–S inequality on the cross product terms to obtain

$$\begin{aligned} &|K(A\mathbf{u}) - \hat{K}(A\mathbf{u})| \\ &\le \sum_{j=1}^{p} \Big\{ |Y_{j\mathbf{u}}^o - Y_{j\mathbf{v}}^o|_n^2 + |\mu_{j\mathbf{u}}^o - \mu_j^o - \mathbf{u}'\boldsymbol{\nu}_j|_n^2 \\ &\quad + 2|Y_{j\mathbf{u}}^o - Y_{j\mathbf{v}}^o|_n \,[|Y_j^o + \mathbf{u}'\boldsymbol{\nu}_j + m_j|_n + |\mu_{j\mathbf{u}}^o - \mu_j^o - \mathbf{u}'\boldsymbol{\nu}_j|_n] \\ &\quad + 2|Y_j^o + \mathbf{u}'\boldsymbol{\nu}_j + m_j|_n \cdot |\mu_{j\mathbf{u}}^o - \mu_j^o - \mathbf{u}'\boldsymbol{\nu}_j|_n \Big\}. \end{aligned} \tag{29}$$

In view of Lemmas 5.5.1, 5.5.2 and assumptons (8) and (10), (28) will follow from (29) if we prove

$$\sup_{\|\mathbf{u}\|\le B} \sum_{j=1}^{p} |\mu_{j\mathbf{u}}^{o} - \mu_{j}^{o} - \mathbf{u}'\nu_j|_n^2 = o(1). \tag{30}$$

Let $\xi_{j\mathbf{u}} := |\mu_{j\mathbf{u}}^{o} - \mu_{j}^{o} - \mathbf{u}'\nu_j|_n^2$, $1 \le j \le p$, $\mathbf{u}\in\mathbb{R}^p$. In view of the compactness of $\mathcal{N}(B)$ and the assumption (9), it suffices to prove that $\forall\ \epsilon > 0$, $\exists$ a $\delta > 0$ $\ni\ \forall\ \mathbf{v} \in \mathcal{N}(B)$,

$$\limsup_n \sup_{\|\mathbf{u}-\mathbf{v}\|\le\delta} \sum_{j=1}^{p} |\xi_{j\mathbf{u}} - \xi_{j\mathbf{v}}| \le \epsilon. \tag{31}$$

But

$$\begin{aligned} |\xi_{j\mathbf{u}} - \xi_{j\mathbf{v}}| \le 2 \Big\{ &|\mu_{j\mathbf{u}}^{o} - \mu_{j\mathbf{v}}^{o}|_n^2 + \|\mathbf{u}-\mathbf{v}\|^2\, \|\nu_j\|_n^2 \\ &+ \xi_{j\mathbf{v}}^{1/2} \big[|\mu_{j\mathbf{u}}^{o} - \mu_{j\mathbf{v}}^{o}|_n + \|\mathbf{u}-\mathbf{v}\|\, \|\nu_j\|_n \big] \\ &+ |\mu_{j\mathbf{u}}^{o} - \mu_{j\mathbf{v}}^{o}|_n\, \|\mathbf{u}-\mathbf{v}\|\, \|\nu_j\|_n \Big\}. \end{aligned}$$

Hence, from (22) and the assumption (9),

$$\text{l.h.s. } (31) \le 2\,\{4k\delta^2 + \delta^2(a + 2k^{1/2}a^{1/2})\} = k_1\delta^2$$

where $a = \limsup_n \sum_{j=1}^{p} \|\nu_j\|_n^2$. Therefore, choose $\delta^2 \le \epsilon/k_1$ to obtain (31), hence (30) and therefore the Theorem. □

Our next goal is to obtain an analogue of (28) for $K_{\mathbf{D}}^{+}$. Before stating it rigorously, it helps to rewrite $K_{\mathbf{D}}^{+}$ in terms of standardized processes $\{Y_j^o\}$ and $\{\mu_j^o\}$ defined at (9). In fact, we have

$$\begin{aligned} K_{\mathbf{D}}^{+}(A\mathbf{u}) &= \sum_{j=1}^{p} \int [S_j^o(y,\mathbf{u}) - \Sigma_i d_{ij} + S_j^o(-y,\mathbf{u})]^2\, dG_n(y) \\ &= \sum_{j=1}^{p} \int [Y_j^o(y,\mathbf{u}) - Y_j^o(y) + Y_j^o(-y,\mathbf{u}) - Y_j^o(-y) \\ &\qquad + \mu_j^o(y,\mathbf{u}) - \mu_j^o(y) - \mathbf{u}'\nu_j(y) \\ &\qquad + \mu_j^o(-y,\mathbf{u}) - \mu_j^o(-y) - \mathbf{u}'\nu_j(y) \\ &\qquad + \mathbf{u}'\nu_j^{+}(y) + W_j^{+}(y) + m_j^{+}(y)]^2\, dG_n(y) \end{aligned}$$

where

$$W_j^+(y) := Y_j^o(y) + Y_j^o(-y), \qquad \nu_j^+(y) = \nu_j(y) + \nu_j(-y),$$

$$m_j^+(y) := \Sigma_i\, d_{ij}\,\{F_i(y) - 1 + F_i(-y)\}$$

$$= \mu_j^o(y) + \mu_j^o(-y) - \Sigma_i\, d_{ij}, \qquad y \in \mathbb{R},\ 1 \le j \le p.$$

Let

$$(32) \qquad \hat{K}_D^+(Au) = \sum_{j=1}^{p} \int [W_j^+ + m_j^+ + u'\nu_j^+]^2\, dG_n, \qquad u \in \mathbb{R}^p.$$

Now proceeding as in (29), one obtains a similar upper bound for $|K_D^+(Au) - \hat{K}_D^+(Au)|$ innvolving terms like those in r.h.s. of (29) and the terms like $|Y_{ju}^o - Y_j^o|_{-n}$, $|\mu_{ju}^o - \mu_j^o - u'\nu_j|_{-n}$ $\|\nu_j\|_{-n}$, $|Y_j|_{-n}$, where for any function $h\colon \mathbb{R}^{p+1} \longrightarrow \mathbb{R}$, $|h_u|_{-n}^2 := \int h^2(-y, u)\, dG_n(y)$. It thus becomes apparent that one needs an analogue of Lemmas 5.5.1 and 5.5.2 with $G_n(\cdot)$ replaced by $G_n(-\cdot)$. That is, if the conditions (5) – (10) are also assumed to hold for measures $\{G_n(-\cdot)\}$ then obviously analogues of these lemmas will hold. Alternatively, the statement of the following theorem and the details of its proof are considerably simplified if one assumes G_n to be symmetric around zero, as we shall do for convenience. Before stating the theorem, we state

Lemma 5.5.3. *Let* $Y_{n1}, \dots, Y_{nn}$ *be independent r.v.'s with respective d.f.'s* $F_{n1}, \dots, F_{nn}$. *Assume* (1) – (4), (6), (7) *hold,* $\{G_n\}$ *satisfies* (5.3.8) *and that* (33) *hold, where*

$$(33) \qquad \int \Sigma_i\, \|d_{ni}\|^2\, \{F_{ni}(-y) + 1 - F_{ni}(y)\}\, dG_n(y) = O(1).$$

Then,

$$(34a) \qquad E \sum_{j=1}^{p} |Y_{j0}^o|_{-n}^2 = O(1),$$

and

$$(34b) \qquad E \sup_{\|u\| \le B} \sum_{j=1}^{p} |Y_{ju}^o - Y_{j0}^o|_{-n}^2 = o(1), \qquad \forall\ 0 < B < \infty. \ \square$$

This lemma follows from Lemmas 5.5.1 and 5.5.2 because under (5.3.8), l.h.s.'s of (34a) and (34b) are equal to those of (14) and (16), respectively. The proof of the following theorem is similar to that of Theorem 5.5.1.

Theorem 5.5.2. *Let* $Y_{n1}, \dots, Y_{nn}$ *be independent r.v.'s with respective d.f.'s* $F_{n1}, \dots, F_{nn}$. *Suppose that* $\{X, F_{ni}, D, G_n\}$ *satisfy* (1) –

(4), (6) – (9), (5.3.8) *for all* $n \geq 1$, (33) *and that*

(35) $$\sum_{j=1}^{p} \int \{m_j^+(y)\}^2 \, dG_n(y) = O(1),$$

Then, $\forall\ 0 < B < \infty$,

(36) $$E \sup_{\|u\| \leq B} |K_D^+(Au) - \hat{K}_D^+(Au)| = o(1). \qquad \square$$

Remark 5.5.1. Recall that we are interested in the asymptotic distribution of $A^{-1}(\hat{\beta}_D - \beta)$ which is a minimizer of $\hat{K}_D(\beta + Au)$ w.r.t. u. Since $\hat{\beta}_D$ satisfies (5.3.22), there is no loss of generality in taking the true β equal to 0. Then (28) asserts that $(1/2)\hat{K}_D$ satisfies (5.4.A1) with

(37) $$\theta_0 = 0,\ \delta_n = A^{-1},\ S_n = A^{-1}\mathcal{T}_n,\ W_n = A\,\mathcal{B}_n\,A,$$

$$\mathcal{T}_n := -\int \Gamma_n(y)\{Y_D^o(y) + m_D(y)\}\, dG_n(y),$$

$$\mathcal{B}_n := \int \Gamma_n(y)\Gamma_n'(y)\, dG_n(y),$$

where $\Gamma_n(y) = AX'\Lambda^*(y)D$, Λ^* as in (4.2.1), $Y_D^{o\prime} := (Y_1^o, \ldots., Y_p^o)$ and $m_D' := (m_1, \ldots., m_p)$.

In view of Lemma 5.5.1, the assumptions (5) and (10) imply that $EK_D(0) = O(1)$, thereby ensuring the validity of (5.4.A4).

Similarly, (36) asserts that $(1/2)K_D^+$ satisfies (5.4.A1) with

(38) $$\theta_0 = 0,\ \delta_n = A^{-1},\ S_n = A^{-1}\mathcal{T}_n^+,\ W_n = A\mathcal{B}_n^+ A,$$

$$\mathcal{T}_n^+ := -\int \Gamma_n^+(y)\{W_D^+(y) + m_D^+(y)\}\, dG_n(y),$$

$$\mathcal{B}_n^+ := \int \Gamma_n^+(y)\Gamma_n^{+\prime}(y)\, dG_n(y),$$

where $\Gamma_n^+(y) := AX'\Lambda^+(y)D$, $\Lambda^+(y) := \Lambda^*(y) + \Lambda^*(-y)$, $y \in \mathbb{R}^p$, $W_D^{+\prime} := (W_1^+, \ldots, W_p^+)$ and $m_D^{+\prime} := (m_1^+, \ldots, m_p^+)$.

In view of (12), (31), (33) and (5.3.8) it follows that (5.4.A4) is satisfied by $K_D^+(0)$.

Theorem 5.4.1 enables one to study the asymptotic distribution of $\hat{\beta}_D$ when in (1.1.1) the actual error d.f. F_{ni} is not necessarily equal to the

modeled d.f. H_{ni}, $1 \le i \le n$. Theorem 5.4.2 enables one to study the asymptotic distribution of β^+_D when in (1.1.1) the error d.f. F_{ni} is not necessarily symmetric around 0, but we model it to be so, $1 \le i \le n$. □

So far we have not used the assumptions (11) and (12). They will be now used to obtain (5.4.A5) for K_D and K^+_D.

Lemma 5.5.4. *In addition to the assumptions of Theorem* 5.5.1 *assume that* (11) *and* (12) *hold. Then,* $\forall\ \epsilon > 0,\ 0 < z < \infty,\ \exists$ N *(depending only on* ϵ*) and a* B *(depending on* ϵ, z*)* $\ni\ 0 < B < \infty$,

$$\text{(39)} \qquad P\Big(\inf_{\|u\|>B} K_D(Au) \ge z\Big) \ge 1-\epsilon, \qquad \forall\ n \ge N,$$

$$\text{(40)} \qquad P\Big(\inf_{\|u\|>B} \hat{K}_D(Au) \ge z\Big) \ge 1-\epsilon, \qquad \forall\ n \ge N.$$

Proof. As usual write K, $\hat{K}$ etc. for K_D, $\hat{K}_D$ etc. Recall the definition of $\bar{\Gamma}_n$ from (11). Let $k_n(\theta) := \theta'\bar{\Gamma}_n\theta$, $\theta \in \mathscr{R}^p$. By the C–S inequality and (11),

$$\text{(41)} \qquad \sup_{\|\theta\|=1} |k_n(\theta)|^2 \le \|\bar{\Gamma}_n\|^2 \le \sum_{j=1}^{p} \|\nu_j\|_n^2 \cdot |g_n|_n^2 = O(1).$$

Fix an $\epsilon > 0$ and a $z \in (0, \infty)$. Define, for $t \in \mathbb{R}^p$, $1 \le j \le p$,

$$\hat{V}_j(t) := \int \{Y_j^o + t'R_j + m_j\}\, g_n\, dG_n,$$

$$V_j(t) := \int [V_{jd}(y, t) - \sum_{i=1}^{n} d_{nij}\, H_{ni}(y)]\, g_n(y)\, dG_n(y).$$

Also, let $\hat{V}' := (\hat{V}_1, \dots, \hat{V}_p)$, $V' := (V_1, \dots, V_p)$, $\gamma_n := |g_n|_n^2$, $\gamma := \limsup_n \gamma_n$.

Write a $u \in \mathbb{R}^p$ with $\|u\| > B$ as $u = r\theta$, $|r| > B$, $\|\theta\| = 1$. Then, by the C–S inequality,

$$\inf_{\|u\|>B} K(Au) \ge \inf_{|r|>B,\|\theta\|=1} (\theta' V(rA\theta))^2/\gamma_n,$$

$$\inf_{\|u\|>B} \hat{K}(Au) \ge \inf_{|r|>B,\|\theta\|=1} (\theta' \hat{V}(rA\theta))^2/\gamma_n.$$

It thus suffices to show that $\exists$ a $B \in (0, \infty)$ and $N \ni$

$$(\tilde{39}) \qquad P\Big(\inf_{|r|>B,\, \|\theta\|=1} (\theta' V(rA\theta))^2/\gamma_n \geq z\Big) \geq 1-\epsilon, \qquad \forall\ n \geq N,$$

$$(\tilde{40}) \qquad P\Big(\inf_{|r|>B,\, \|\theta\|=1} (\theta' \hat{V}(rA\theta))^2/\gamma_n \geq z\Big) \geq 1-\epsilon, \qquad \forall\ n \geq N.$$

But, $\forall\ \mathbf{u}\in\mathbb{R}^p$,

$$\|V(A\mathbf{u}) - \hat{V}(A\mathbf{u})\| \leq 2\gamma_n \sum_{j=1}^{p} \{|Y_{j\mathbf{u}}^o - Y_{j\mathbf{v}}^o|_n^2 + |\mu_{j\mathbf{u}}^o - \mu_j^o - \mathbf{u}'\nu_j|_n^2\}.$$

Thus, from (11), (16) and (30), it follows that $\forall\ B\in(0,\infty)$,

$$(42) \qquad \sup_{\|\mathbf{u}\|\leq B} \|V(A\mathbf{u}) - \hat{V}(A\mathbf{u})\| = o_p(1).$$

Now rewrite

$$\theta' \hat{V}(rA\theta) = \theta' T + r\, k_n(\theta), \qquad T' := (T_1, \ldots, T_p) \quad \text{with}$$

$$T_j := \int \{Y_j^o + m_j\}\, g_n\, dG_n, \qquad 1 \leq j \leq p.$$

Again, by the C–S inequality, Fubini, (16) and the assumptions (10) and (11) it follows that $\exists\ N_1$ and b, possibly both depending on ϵ, such that

$$(43) \qquad P(\|T\| \leq b) \geq 1 - (\epsilon/2), \qquad n \geq N_1.$$

Now choose B such that

$$(44) \qquad B \geq (b + (za)^{1/2})\, \alpha^{-1},$$

where α is as in (11). Then, with $\alpha_n := \inf\{|k_n(\theta)|;\ \|\theta\| = 1\}$,

$$\begin{aligned}
(45) \quad & P\Big(\inf_{|r|=B,\, \|\theta\|=1} (\theta' \hat{V}(rA\theta))^2/\gamma_n \geq z\Big) \\
&= P(|\theta' \hat{V}(rA\theta)| \geq (z\gamma_n)^{1/2},\ \forall\ \|\theta\| = 1,\ |r| = B) \\
&\geq P(||\theta' T| - |r|\ |k_n(\theta)|| \geq (z\gamma_n)^{1/2},\ \forall\ \|\theta\| = 1,\ |r| = B) \\
&\geq P(\|T\| \leq -(z\gamma_n)^{1/2} + B\,\alpha_n) \geq P(\|T\| \leq -(z\gamma)^{1/2} + B\,\alpha) \\
&\geq P(\|T\| \leq b) \geq 1 - (\epsilon/2), \qquad \forall\ n \geq N_1.
\end{aligned}$$

In the above, the first inequality follows from the fact that $||d| - |c|| \le |d + c|$, d, c reals; the second uses the fact that $|\boldsymbol{\theta}' \boldsymbol{T}| \le \|\boldsymbol{T}\|$ for all $\|\boldsymbol{\theta}\| = 1$; the third uses the relation $(-\infty, -(z\gamma)^{1/2} + B\,\alpha) \subset (-\infty, -(z\gamma_n)^{1/2} + B\,\alpha_n)$; while the last inequality follows from (43) and (44).

Observe that $\boldsymbol{\theta}' \hat{\boldsymbol{V}}(rA\boldsymbol{\theta})$ is monotonic in r for every $\|\boldsymbol{\theta}\| = 1$. Therefore, (45) implies $(4\tilde{0})$ and hence (40) in a straight forward fashion.

Next, consider $\boldsymbol{\theta}' \boldsymbol{V}(rA\boldsymbol{\theta})$. Rewrite

$$\boldsymbol{\theta}' \boldsymbol{V}(rA\boldsymbol{\theta}) = \int \sum_{i=1}^{n} (\boldsymbol{\theta}' d_{ij})[I(Y_{ni} \le y + r x_{ni}' A\boldsymbol{\theta})) - H_{ni}(y)]\, g_n(y)\, dG_n(y)$$

which, in view of the assumption (12), shows that $\boldsymbol{\theta}' \boldsymbol{V}(rA\boldsymbol{\theta})$ is monotonic in r for every $\|\boldsymbol{\theta}\| = 1$. Therefore, by (42) $\exists$ N_2, depending on ϵ, $\ni$

$$\begin{aligned}
&P\Big(\inf_{|r|>B,\,\|\boldsymbol{\theta}\|=1} (\boldsymbol{\theta}' \boldsymbol{V}(rA\boldsymbol{\theta}))^2/\gamma_n \ge z\Big) \\
&\qquad \ge P\Big(\inf_{|r|=B,\,\|\boldsymbol{\theta}\|=1} (\boldsymbol{\theta}' \boldsymbol{V}(rA\boldsymbol{\theta}))^2/\gamma_n \ge z\Big) \\
&\qquad \ge P\Big(\inf_{|r|=B,\,\|\boldsymbol{\theta}\|=1} (\boldsymbol{\theta}' \hat{\boldsymbol{V}}(rA\boldsymbol{\theta}))^2/\gamma_n \ge z\Big) - (\epsilon/2), && \forall\, n \ge N_2, \\
&\qquad \ge 1 - \epsilon, && \forall\, n \ge N_2 \vee N_1,
\end{aligned}$$

by (45). This proves $(3\tilde{9})$ and hence (39). □

The next lemma gives an analogue of the previous lemma for $K_{\mathbf{D}}^{+}$. Since the proof is quite similar no details will be given.

Lemma 5.5.5. *In addition to the assumptions of Theorem* 5.5.2 *assume that* (11^{+}) *and* (12) *hold, where* (11^{+}) *is the condition* (11) *with* Γ_n *replaced by* $\Gamma_n^{+} := (\nu_1^{+}, \dots, \nu_p^{+})$ *and where* $\{\nu_j^{+}\}$ *are defined just above* (32).

Then, $\forall\ \epsilon > 0,\ 0 < z < \infty,\ \exists$ N *(depending only on* ϵ*) and a* B *(depending on* ϵ, z*)* $\ni$

$$P\Big(\inf_{\|u\|>B} K_{\mathbf{D}}^{+}(Au) \ge z\Big) \ge 1-\epsilon, \qquad \forall\ n \ge N, \tag{46}$$

$$P\Big(\inf_{\|u\|>B} \hat{K}_{\mathbf{D}}^{+}(Au) \ge z\Big) \ge 1-\epsilon, \qquad \forall\ n \ge N. \ \square \tag{47}$$

The above two lemmas verify (5.4.A5) for the two dispersions K and K^+. Also note that (40) together with (5) and (10) imply that $\|A^{-1}(\hat{\Delta} - \beta)\| = O_p(1)$, where $\hat{\Delta}$ is defined at (49) below. Similarly, (47), (5), (35) and the symmetry assumption (5.3.8) about $\{G_n\}$ imply that $\|A^{-1}(\Delta^+ - \beta)\| = O_p(1)$, where Δ^+ is defined at (53) below. The proofs of these facts are exactly similar to that of (5.4.2) given in the proof of Theorem 5.4.1.

In view of Remark 5.5.1 and Theorem 5.4.1, we have now proved the following theorems.

Theorem 5.5.3. *Assume that* (1.1.1) *holds with the modeled and actual d.f.'s of the errors* $\{e_{ni}, 1 \le i \le n\}$ *equal to* $\{H_{ni}, 1 \le i \le n\}$ *and* $\{F_{ni}, 1 \le i \le n\}$, *respectively. In addition, suppose that* (1) – (12) *hold. Then*

$$(48) \qquad (\hat{\beta}_D - \hat{\Delta})' A^{-1} \mathscr{B}_n A^{-1}(\hat{\beta}_D - \hat{\Delta}) = o_p(1),$$

where $\hat{\Delta}$ *satisfies the equation*

$$(49) \qquad \mathscr{B}_n A^{-1}(\hat{\Delta} - \beta) = \mathscr{T}_n.$$

If, in addition,

$$(50) \qquad \mathscr{B}_n^{-1} \text{ exists for } n \ge p,$$

then,

$$(51) \qquad A^{-1}(\hat{\beta}_D - \beta) = \mathscr{B}_n^{-1} \mathscr{T}_n + o_p(1),$$

where $\mathscr{T}_n$ *and* $\mathscr{B}_n$ *are defined at* (37). □

Theorem 5.5.4. *Assume that* (1.1.1) *holds with the actual d.f.'s of the errors* $\{e_{ni}, 1 \le i \le n\}$ *equal to* $\{F_{ni}, 1 \le i \le n\}$. *In addition, suppose that* $\{X, F_{ni}, D, G_n\}$ *satisfy* (1)–(4), (6) – (9), (5.3.8) *for all* $n \ge 1$, (11), (12) *and* (33). *Then,*

$$(52) \qquad (\beta_D^+ - \Delta^+)' A^{-1} \mathscr{B}_n^+ A^{-1}(\beta_D^+ - \Delta^+) = o_p(1),$$

where Δ^+ *satisfies the equation*

$$(53) \qquad \mathscr{B}_n^+ A^{-1}(\Delta^+ - \beta) = \mathscr{T}_n^+.$$

If, in addition,

$$(54) \qquad (\mathscr{B}_n^+)^{-1} \text{ exists for } n \ge p,$$

then,

$$A^{-1}(\beta_D^+ - \beta) = (\mathscr{B}_n^+)^{-1}\mathscr{S}_n^+ + o_p(1), \tag{55}$$

where $\mathscr{S}_n^+$ *and* $\mathscr{B}_n^+$ *are defined at* (38). □

Remark 5.5.2. If $\{F_i\}$ are symmetric about zero then $\mathbf{m}_D^+ \equiv \mathbf{0}$ and β_D^+ is consistent for β even if the errors are not identically distributed. On the other hand, if the errors are identically distributed, but not symmetrically, then β_D^+ will be asymptotically biased. This is not surprising because here the symmetry, rather than the identically distributed nature of the errors is relevant.

If $\{F_i\}$ are symmetric about an unknown common point then that point can be also estimated by the above m.d. method by simply augmenting the design matrix to include the column 1, if not present already. □

Next we turn to the K_D^* and β_D^* (5.2.22) and (5.2.23). First we state a theorem giving an analogue of (28) for K_D^*. Let $\mathcal{Y}_j$, μ_j be $\mathcal{Y}_d$, μ_d of (2.3.1) with $\{d_{ni}\}$ replaced by $\{d_{nij}\}$, $j = 1, \dots, p$, X_{ni} replaced by Y_{ni} and $c_{ni} = A_1(\mathbf{x}_{ni} - \bar{\mathbf{x}}_n)$, $1 \le i \le n$, where A_1 and $\bar{\mathbf{x}}_n$ are defined at (4.3.11). Set

$$\mathbf{R}_j^*(s) := \Sigma_i\, (d_{nij} - \tilde{d}_{nj}(s))(\mathbf{x}_{ni} - \bar{\mathbf{x}}_n)\, q_{ni}(s), \tag{56}$$

where, for $1 \le j \le p$, $\tilde{d}_{nj}(s) := n^{-1}\,\Sigma_i\, d_{nij}\,\ell_{ni}(s)$, $0 \le s \le 1$, with $\{\ell_{ni}\}$ as in (3.2.35) and $q_{ni} \equiv f_{ni}(H^{-1})$, $1 \le i \le n$. Let

$$\hat{K}_D^*(\mathbf{t}) := \sum_{j=1}^{p} \int_0^1 \{\mathcal{Y}_j(s, 0) - \mathbf{t}'\mathbf{R}_j^*(s) + \mu_j(s, 0)\}^2\, dL_n(s). \tag{57}$$

In (59) below, L in K_D^* is supposed to have been replaced by L_n.

Theorem 5.5.5. *Let* $Y_{n1}, \dots, Y_{nn}$ *be independent r.v.'s with respective d.f.'s* $F_{n1}, \dots, F_{nn}$. *Assume* $\{D, X, F_{ni}\}$ *satisfy* (1), (2), (3), (2.3.3b), (3.2.12), (3.2.35) *and* (3.2.36) *with* $w_i = d_{ij}$, $1 \le j \le p$, $1 \le i \le n$. *Let* $\{L_n\}$ *be a sequence of d.f.'s on* [0, 1] *and assume that*

$$\sum_{j=1}^{p} \int_0^1 \mu_j^2(s, 0)\, dL_n(s) = O(1). \tag{58}$$

Then, for every $0 < B < \infty$,

$$\sup_{\|\mathbf{u}\| \le B} |K_D^*(A\mathbf{u}) - \hat{K}_D^*(A\mathbf{u})| = o_p(1). \tag{60}$$

Proof. The proof of (60) uses the a.u.l. result of Theorems 3.2.1 and 3.2.4. Details are left out as an exercise. □

The result (60) shows that the dispersion K_D^* satisfies (5.4.A1) with

$$\theta_0 = 0,\ \delta_n = A_1^{-1},\ S_n = A_1^{-1}\mathcal{S}_n^*,\ W_n = A_1\mathcal{B}_n^* A_1, \tag{61}$$

$$\mathcal{S}_n^* := -\int_0^1 \Gamma_n^*(s)\{Y_D(s) + \mu_D(s)\}\, dL_n(s),$$

$$\mathcal{B}_n^* := \int_0^1 \Gamma_n^*(s)\Gamma_n^{*\prime}(s)\, dL_n(s),$$

where $\Gamma_n^*(s) = A_1'X_c\Lambda(s)D(s)$, $D(s) := ((d_{nij} - \tilde{d}_{nj}(s)),\ 1\le i\le n,\ 1\le j\le p;\ \Lambda(s)$ as in (2.3.32), $0 \le s \le 1$; X_c as in (4.2.11); $Y_D := (Y_1, \ldots, Y_p)'$, $\mu_D' := (\mu_1, \ldots, \mu_p)$ *with* $Y_j(s) \equiv Y_j(s, 0)$, $\mu_j(s) \equiv \mu_j(s, 0)$.

Call the condition (11) by the name of (11*) if it holds when (Γ_n, G_n) is replaced by (Γ_n^*, L_n). Analogous to Theorem 5.5.4 we have

Theorem 5.5.6. *Assume that* (1.1.1) *holds with the actual d.f.'s of the errors* $\{e_{ni},\ 1 \le i \le n\}$ *equal to* $\{F_{ni},\ 1 \le i \le n\}$. *In addition, assume that* $\{D, X, F_{ni}\}$ *satisfy* (NX^*), (2), (3), (2.3.3b), (3.2.12), (3.2.35), (3.2.36) *with* $w_i = d_{ij}$, $1 \le j \le p$, $1 \le i \le n$, (11*) *and* (12). *Let* $\{L_n\}$ *be a sequence of d.f.'s on* $[0, 1]$ *satisfying* (58). *Then*

$$(\beta_D^* - \Delta^*)' A^{-1}\mathcal{B}_n^* A^{-1}(\beta_D^* - \Delta^*) = o_p(1), \tag{62}$$

where Δ^* *satisfies the equation*

$$\mathcal{B}_n^* A^{-1}(\Delta^* - \beta) = \mathcal{S}_n^*. \tag{63}$$

If, in addition,

$$(\mathcal{B}_n^*)^{-1} \text{ exists for } n \ge p, \tag{64}$$

then,

$$A^{-1}(\beta_D^* - \beta) = (\mathcal{B}_n^*)^{-1}\mathcal{S}_n^* + o_p(1), \tag{65}$$

where $\mathcal{S}_n^*$ *and* $\mathcal{B}_n^*$ are *defined at* (61).

The proof of this theorem is similar to that of Theorem 5.5.3. The details are left out for interested readers. See also Section 4.3. □

Remark 5.5.3. *Discussion of the assumptions* (1) – (10). Among the assumptions (1) – (10), the assumptions (7) and (9) are relatively harder to verify. First, we shall give some sufficient conditions that will imply (7), (9) and the other assumptions. Then, we shall discuss these assumptions in detail for three cases, v.i.z., the case when the errors are correctly modeled to be i.i.d. F, F a known d.f., the case when we model the errors to be i.i.d. F but they actually have heteroscedastic gross errors distributions, and finally, the case when the errors are modeled to be i.i.d. F but they actually are heteroscedastic due to difference in scales.

To begin with consider the following assumptions.

(66) *For any sequences of numbers* $\{a_{ni}, b_{ni}\}$, $a_{ni} < b_{ni}$, $\max_i (b_{ni} - a_{ni}) \longrightarrow 0$,

$$\limsup_n \max_i (b_{ni}-a_{ni})^{-1} \int_{a_{ni}}^{b_{ni}} \int \{f_{ni}(y+z)-f_{ni}(y)\}^2\, dG_n(y)\, dz = 0.$$

(67) $$\max_i \int f_{ni}^2\, dG_n = O(1).$$

Claim 5.5.1. *Assumptions* (1) – (4), (66), (67) *imply* (7) *and* (9).

Proof. Use the C–S inequality twice, the fact that $(d_{ij}^{\pm})^2 \le d_{ij}^2$ for all i, j, and (2) to obtain

$$\sum_{j=1}^{p} \int [\sum_i d_{ij}^{\pm} \{F_i (y + c_i'v + \delta\kappa_i) - F_i (y + c_i'v - \delta\kappa_i)\}]^2\, dG_n(y)$$

$$\le 2\, \Sigma_i \|d_i\|^2 \int \Sigma_i\, \delta\kappa_i \int_{a_i}^{b_i} f_i^2 (y + z)\, dz\, dG_n(y)$$

$$\le 4p^2\delta^2 \max_i (2\delta\kappa_i)^{-1} \int_{a_i}^{b_i} \int f_i^2 (y+z)\, dG_n(y) dz, \quad \text{(by Fubini)},$$

where $a_i = -\kappa_i\delta + c_i'v$, $b_i = \kappa_i\delta + c_i'v$, $1 \le i \le n$. Therefore, by (66), (67) and (1),

$$\text{l.h.s. } (7) \le 4p^2\delta^2 k, \qquad (k = \limsup_n \max_i |f_i|_n^2),$$

which shows that (7) holds.

Next, by (2) and two applications of the C–S inequality

$$\text{l.h.s. } (9) = \sum_{j=1}^{p} \int [\Sigma_i\, d_{ij} \{F_i (y + c_i'u) - F_i(y) - c_i'u f_i(y)\}]^2\, dG_n(y)$$

$$\le p \int \Sigma_i \{F_i(y + c_i'u) - F_i(y) - c_i'u f_i(y)\}^2\, dG_n(y)$$

$$= p\Big\{\int \Sigma_i^+ \Big[\int_0^{c_i'u} (f_i(y+z) - f_i(y))dz\Big]^2 dG_n(y)$$

$$+ \int \Sigma_i^- \Big[-\int_{c_i'u}^0 (f_i(y+z) - f_i(y))dz\Big]^2 dG_n(y)\Big\}$$

$$\le 2p\,\Big\{\int [\Sigma_i^+ c_i'u \int_0^{c_i'u} \{f_i(y+z) - f_i(y)\}^2\, dz$$

$$+ \Sigma_i^- (-c_i'u) \int_{c_i'u}^0 \int \{f_i(y+z) - f_i(y)\}^2\, dz]\, dG_n(y)$$

$$\le [\max_i (2|c_i'u|)^{-1} \int_{-|c_i'u|}^{|c_i'u|} \int \{f_i(y+z) - f_i(y)\}^2 dG_n(y)dz] \cdot$$

$$\cdot 4p\, \Sigma_i (c_i'u)^2,$$

where Σ_i^+ (Σ_i^-) is the sum over those i for which $c_i'u \ge 0$ $(c_i'u < 0)$. Since $\Sigma_i (c_i'u)^2 \le pB$ for all $u \in \mathcal{N}(B)$, (9) now follows from (66) and (1). □

Now we consider the three special cases mentioned above.

***Case* 5.5.1.** *Correctly modeled i.i.d. errors*: $F_{ni} \equiv F \equiv H_{ni}$, $G_n \equiv G$. Suppose that F has a density f w.r.t. λ. Assume that

(68) (a) $0 < \int f\, dG < \infty$, (b) $0 < \int f^2\, dG < \infty$.

(69) $\int F(1-F)\, dG < \infty$.

(70) (a) $\lim_{z\to 0} \int f(y+z)\, dG(y) = \int f\, dG$

(b) $\lim_{z\to 0} \int f^2(y+z)\, dG(y) = \int f^2\, dG$.

***Claim* 5.5.2:** *Assumptions* (1), (2), (4) *with* $G_n \equiv G$, (68) – (70) *imply* (1) – (10) *with* $G_n \equiv G$.

This is easy to see. In fact here (5) and (6) are equivalent to (68a), (69) and (70a); (66) and (67) are equivalent to (68b) and (70b). The LHS (10) = 0.

Note that if G is absolutely continuous then (68) implies (70). If G is purely discrete and f continuous at the points of jumps of G then (70) holds. In particular if $G = \delta_0$, i.e., if G is degenerate at 0, $\infty > f(0) > 0$ and f is continuous at 0 then (68), (70) are trivially satisfied. If $G(y) \equiv y$, (68a) and (70a) are *a priori* satisfied while (69) is equivalent to assuming that $E|e_1 - e_2| < \infty$, e_1, e_2 i.i.d. F.

If $dG(y) = \{F(y)(1 - F(y))\}^{-1} dF(y)$, the so called Darling–Anderson measure, then (68) – (70) are satisfied by a class of d.f.'s that includes normal, logistic and double exponential distributions.

Case 5.5.2. *Heteroscedastic gross errors*: $H_{ni} \equiv F$, $F_{ni} \equiv (1-\delta_{ni})F + \delta_{ni} F_0$. We shall also assume that $G_n \equiv G$. Let f and f_0 be continuous densities of F and F_0. Then $\{F_{ni}\}$ have densities $f_{ni} = f + \delta_{ni}(f_0 - f)$, $1 \le i \le n$. Hence (3) is satisfied. Consider the assumption

$$0 \le \delta_{ni} \le 1, \qquad \max_i \delta_{ni} \to 0, \tag{71}$$

$$\int |F_0 - F| \, dG < \infty. \tag{72}$$

Claim 5.5.3. *Suppose that* f_0 *and* f *satisfy* (68) *and* (70), F *satisfies* (69), *and suppose that* (1), (2) *and* (4) *hold. Then* (71) *and* (72) *imply* (5) – (9).

Proof. The relation $f_i \equiv f + \delta_i(f_0 - f)$ implies that

$$\nu_j - \Sigma_i d_{ij} c_i f = \Sigma_i d_{ij} c_i \delta_i (f_0 - f), \qquad 1 \le j \le p,$$

and

$$\gamma_n - \Sigma_i \|d_i\|^2 f = \Sigma_i \|d_i\|^2 \delta_i (f_0 - f).$$

Because $\Sigma_i \|d_i\|^2 \le p$, $\Sigma_i \|c_i\|^2 = p$, we obtain

$$\begin{aligned} &\left| \int [\gamma_n(y+x) - \Sigma_i \|d_i\|^2 f(y + x)] dG(y) \right| \\ &\qquad \le p \max_i \delta_i \left| \int [f_0(y + x) - f(y + x)] dG(y) \right|, \qquad \forall\ x \in \mathbb{R}. \end{aligned}$$

Therefore, by (71), (68a) and (70a), it follows that (6) is satisfied. Similarly, the inequality

$$\sum_{j=1}^{p} \int \|\nu_j - \Sigma_i d_{ij} c_i f\|^2 \, dG \le 2p^2 \max_i \delta_i^2 \left\{ \int f_0^2 \, dG + \int f^2 \, dG \right\}$$

ensures the satisfaction of (8). The inequality

$$\left| \int \Sigma_i \|d_i\|^2 \{F_i(1 - F_i) - F(1 - F)\} \, dG \right| \le 2p \max_i \delta_i \int |F_0 - F| \, dG,$$

(69), (71) and (72) imply (5). Next,

$$\int \{f_i(y+x) - f_i(y)\}^2 \, dG(y)$$

$$\leq 2(1+2\delta_i^2) \int \{f(y+x) - f(y)\}^2 \, dG(y) + 4\delta_i^2 \int \{f_0(y+x) - f_0(y)\}^2 \, dG(y).$$

Note that (68b), (70b) and the continuity of f imply that

$$\lim_{x\to 0} \int \{f(y+x) - f(y)\}^2 \, dG(y) = 0$$

and a similar result for f_0. Therefore from the above inequality, (70) and (71) we see that (66) and (67) are satisfied. By Claim 5.5.1, it follows that (7) and (9) are satisfied. □

Suppose that G is a *finite measure*. Then (F1) implies (68) – (70) and (72). In particular these assumptions are satisfied by all those f's that have finite Fisher information.

The assumption (10), in view of (72), amounts to requiring that

(73) $$\sum_{j=1}^{p} (\sum_i d_{ij}\, \delta_i)^2 = O(1).$$

But

(74) $$\sum_{j=1}^{p} (\sum_i d_{ij} \delta_i)^2 = \sum_{i=1}^{n} \sum_{k=1}^{n} d_i' \, \delta_i \, d_k \, \delta_k \leq (\sum_i \|d_i\| \, \delta_i)^2.$$

This and (2) suggest a choice of $\delta_i \equiv p^{-1/2} \|d_i\|$ will satisfy (73). Note that if $D = XA$ then $\|d_i\|^2 \equiv \mathbf{x}_i'(X'X)^{-1}\mathbf{x}_i$.

When studying the robustness of $\hat{\beta}_X$ in the following section, $\delta_i^2 \equiv p^{-1}\mathbf{x}_i'(X'X)^{-1}\mathbf{x}_i$ is a natural choice to use. It is an analogue of $n^{-1/2}$–contamination in the i.i.d. setup. □

***Case* 5.5.3.** *Heteroscedastic scale errors*: $H_{ni} \equiv F$, $F_{ni}(y) \equiv F(\tau_{ni} y)$, $G_n \equiv G$. Let F have continuous density f. Consider the conditions

(75) $$\tau_{ni} \equiv \sigma_{ni} + 1; \quad \sigma_{ni} > 0, \; 1 \leq i \leq n; \quad \max_i \sigma_{ni} \longrightarrow 0.$$

(76) $$\lim_{s\to 1} \int |y|^j f^k(sy) \, dG(y) = \int |y|^j f^k(y) \, dG(y), \; j = 1, k = 1;$$

$$j = 0, k = 1, 2.$$

***Claim* 5.5.4.** *Under* (1), (2), (4) *with* $G_n \equiv G$, (68) – (70), (75) *and* (76), *the assumptions* (5) – (9) *are satisfied.*

Proof. By (41), (43), (49) and Theorems II.4.2.1 and V.1.3.1 of Hájek–Sĭdák (op. cit.),

$$(77)\qquad \lim_{x\to 0} \limsup \max_i \int |f(\tau_i(y+x)) - f(y+x)|^r \, dG(y) = 0,$$
$$\lim_{x\to 0} \int |f(y+x) - f(y)|^r \, dG(y) = 0, \qquad r = 1, 2.$$

Now,

$$|\int \Sigma_i \|d_i\|^2 \{F_i(1-F_i) - F(1-F)\} \, dG|$$
$$\le 2p\max_i \int |F(\tau_i y) - F(y)| \, dG(y) \le 2p \max_i \int_1^{\tau_i} \int |y| \, f(sy) \, dG(y) \, ds$$
$$= o(1), \qquad \text{by (48) and (49) with } j = 1, r = 1.$$

Hence (69) implies (5). Next,

$$|\int \gamma_n(y+x) \, dG(y) - \Sigma_i \|d_i\|^2 \int f \, dG|$$
$$\le \Sigma_i \|d_i\|^2 \tau_i \int \{|f(\tau_i(y+x)) - f(y+x)| + |f(y+x) - f(y)|\} dG(y) +$$
$$+ \max_i \sigma_i \, p \int f dG.$$

Therefore, in view of (48), (77) and (68) we obtain (6). Next, consider

$$\int \{f_i(y+x) - f_i(y)\}^2 \, dG(y)$$
$$\le 4\tau_i^2 \int \Big\{[f(\tau_i(y+x)) - f(y+x)]^2 + [f(y+x) - f(y)]^2 +$$
$$+ [f(\tau_i y) - f(y)]^2\Big\} \, dG(y)$$

Therefore, (75) and (50) imply (39), and hence (7) and (9) by Claim 5.5.1. Note that (50) and (41b) imply (40). Finally,

$$\sum_{j=1}^{p} \int \|\nu_j - \sum_i d_{ij} \, c_i f\|^2 \, dG$$
$$\le p^2 \max_i \int \{\tau_i f(\tau_i y) - f(y)\}^2 \, dG(y)$$
$$\le 2p^2 \max_i \tau_i^2 \Big[\int \{f(\tau_i y) - f(y)\}^2 \, dG(y) + \int f^2 \, dG\Big] = o(1),$$

by (75), (70b), (77). Hence (70b) and the fact that $\sum_{j=1}^{p} \|\sum_i d_{ij} \, c_i\|^2 \le p^2$ implies (8). □

Here, the assumption (10) is equivalent to having

$$\sum_{j=1}^{p} \int [\sum_i d_{ij}\{F(\tau_i y) - F(y)\}]^2 \, dG(y) = O(1). \tag{78}$$

One sufficient condition for (78), besides requiring F to have density f satisfying

$$\lim_{s \to 1} \int (yf(sy))^2 \, dG(y) = \int (yf(y))^2 \, dG(y) < \infty, \tag{79}$$

is to have

$$\sum_{i=1}^{n} \sigma_i^2 = O(1). \tag{80}$$

One choice of $\{\sigma_i\}$ satisfying (80) is $\sigma_i^2 \equiv n^{-1/2}$ and the other choice is $\sigma_i^2 \propto \mathbf{x}_i'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i$, $1 \le i \le n$.

Again, if f satisfies **(F1)**, **(F3)** and G is a finite measure then (68) (70), (76) and (79) are *a priori* satisfied. □

Now we shall give a set of sufficient conditions that will yield (5.4.A1) for the Q of (5.2.13). Since Q does not satisfy (5.3.21), the distribution of Q under (1.1.1) is not idependent of $\boldsymbol{\beta}$. Therefore care has to be taken to exhibit this dependence clearly when formulating a theorem pertaining to Q. This of course complicates the presentation somewhat. As before with $\{H_{ni}\}$, $\{F_{ni}\}$ denoting the modeled and the actual d.f.'s of $\{e_{ni}\}$, define for $0 \le s \le 1$, $y \in \mathbb{R}$, $\mathbf{t} \in \mathbb{R}^p$,

$$\begin{aligned} \bar{H}_n(s, y, \mathbf{t}) &:= n^{-1} \sum_{i=1}^{ns} H_{ni}(y - \mathbf{x}_{ni}'\mathbf{t}), \\ m_n(s, y) &:= n^{-1/2} \sum_{i=1}^{ns} \{F_{ni}(y - \mathbf{x}_{ni}'\boldsymbol{\beta}) - H_{ni}(y - \mathbf{x}_{ni}'\boldsymbol{\beta})\}, \\ \mathcal{M}_{1n}(s, y) &:= n^{-1/2} \sum_{i=1}^{ns} \{I(Y_{ni} \le y) - F_{ni}(y - \mathbf{x}_{ni}'\boldsymbol{\beta})\}, \\ d\alpha_n(s, y) &:= dL_n(s) \, dG_n(y). \end{aligned} \tag{81}$$

Observe that

$$Q(\mathbf{t}) = \int [\mathcal{M}_{1n}(s, y) + m_n(s, y) - n^{1/2}\{\bar{H}_n(s, y, \mathbf{t}) - \bar{H}_n(s, y, \boldsymbol{\beta})\}]^2 \, d\alpha_n(s, y).$$

Note that the single integral is over the set $[0, 1] \times \mathbb{R}$.

Assume that $\{H_{ni}\}$ have densities $\{h_{ni}\}$ w.r.t λ and set

$$\text{(82)} \qquad \bar{\mathbf{R}}_n(s, y) := n^{-1/2} \sum_{i=1}^{ns} \mathbf{x}_{ni}\, h_{ni}(y - \mathbf{x}_{ni}'\boldsymbol{\beta}),$$

$$\overline{h_n^2}(y) := n^{-1} \sum_{i=1}^{n} h_{ni}^2(y - \mathbf{x}_{ni}'\boldsymbol{\beta}), \qquad s \in [0, 1],\ y \in \mathbb{R},$$

$$\bar{\boldsymbol{\nu}}_n := \mathbf{A}\bar{\mathbf{R}}_n, \qquad \mathscr{B}_{1n} := \int \bar{\boldsymbol{\nu}}_n \bar{\boldsymbol{\nu}}_n'\, d\alpha_n.$$

Finally define, for $\mathbf{t} \in \mathbb{R}^p$,

$$\text{(83)} \qquad \hat{Q}(\mathbf{t}) := \int [M_{1n}(s, y) + m_n(s, y) + \mathbf{t}'\bar{\mathbf{R}}_n(s, y)]^2\, d\alpha_n(s, y).$$

Theorem 5.5.7. *Assume that* (1.1.1) *holds with the actual and the modeled d.f.'s of the errors* $\{e_{ni}, 1 \le i \le n\}$ *equal to* $\{F_{ni}, 1 \le i \le n\}$ *and* $\{H_{ni}, 1 \le i \le n\}$, *respectively. In addition, assume that* (1) *holds,* $\{H_{ni}, 1 \le i \le n\}$ *have densities* $\{h_{ni}, 1 \le i \le n\}$ *w.r.t.* λ, *and the following hold.*

$$\text{(84)} \qquad |\overline{h_n^2}|_n = O(1).$$

(85) $\forall\ \mathbf{v} \in \mathscr{N}(B), \forall\ \delta > 0,$

$$\limsup_n \max_i (2\delta\kappa_{ni})^{-1} \int_{a_{ni}}^{b_{ni}} \int h_{ni}^2(y - \mathbf{x}_{ni}'\boldsymbol{\beta} + z)\, dG_n(y)\, dz$$

$$= \limsup_n \max_i \int h_{ni}^2(y - \mathbf{x}_{ni}'\boldsymbol{\beta})\, dG_n(y) < \infty,$$

where $a_{ni} = -\delta\kappa_{ni} - \mathbf{c}_{ni}'\mathbf{v}$, $b_{ni} = \delta\kappa_{ni} - \mathbf{c}_{ni}'\mathbf{v}$, $\kappa_{ni} = \|\mathbf{c}_{ni}\|$, $\mathbf{c}_{ni} = \mathbf{A}\mathbf{x}_{ni}$, $1 \le i \le n$.

(86) $\forall\ \mathbf{u} \in \mathscr{N}(B),$

$$\int \{n^{1/2}[\bar{H}_n(s, y, \boldsymbol{\beta} + \mathbf{A}\mathbf{u}) - \bar{H}_n(s, y, \boldsymbol{\beta})] + \mathbf{u}'\bar{\boldsymbol{\nu}}_n\}^2\, d\alpha_n(s, y) = o(1).$$

$$\text{(87)} \qquad \int n^{-1} \sum_{i=1}^{n} F_{ni}(y - \mathbf{x}_{ni}'\boldsymbol{\beta})\, (1 - F_{ni}(y - \mathbf{x}_{ni}'\boldsymbol{\beta}))\, dG_n(y) = O(1).$$

$$\text{(88)} \qquad \int m_n^2(s, y)\, d\alpha_n(s, y) = O(1).$$

Then, $\forall\ 0 < B < \infty,$

$$\text{(89)} \qquad E \sup_{\|\mathbf{u}\| \le B} |Q(\boldsymbol{\beta} + \mathbf{A}\mathbf{u}) - \hat{Q}(\mathbf{A}\mathbf{u})| = o(1).$$

The details of the proof are similar to those of Theorem 5.5.1 and are left out as an exercise for interested readers.

An analogue of (51) for $\bar{\beta}$ will appear in the next section as Theorem 5.6a.3. Its asymptotic distribution in the case when the errors are correctly modeled to be i.i.d. will be also discussed there.

We shall end this section by stating analogues of some of the above results that will be useful when an unknown scale is also being estimated. To begin with, consider K_D of (5.2.24). To simplify writing, let

$$K_D^o(s, u) := K_D((1+sn^{-1/2}), Au), \qquad s \in \mathbb{R},\ u \in \mathbb{R}^p. \tag{90}$$

Write $a_s := (1 + sn^{-1/2})$. Then from (5.2.24) and (90),

$$K_D^o(s, u) = \sum_{j=1}^{p} \int \{Y_j^o(ya_s, u) + \mu_j^o(ya_s, u) - \sum_i d_{ij}\, H_i(y)\}^2 dG_n(y) \tag{91}$$

where H_i is the d.f. of e_i, $1 \le i \le n$, and where μ_j^o, Y_j^o are as in (9) and (13), respectively. Writing $\mu_j^o(y)$, $Y_j^o(y)$ etc. for $\mu_j^o(y, 0)$, $Y_j^o(y, 0)$ etc., we obtain

$$\begin{aligned} K_D^o(s, u) = \sum_{j=1}^{p} \int \{ & Y_j^o(ya_s, u) - Y_j^o(y) + \mu_j^o(ya_s) - \mu_j^o(y) - sy\nu_j^*(y) \\ & + Y_j^o(y) + u'\nu_j(y) + sy\nu_j^*(y) + m_j(y) \\ & + \mu_j^o(ya_s, u) - \mu_j^o(ya_s) - \mu'\nu_j(ya_s) \\ & + u'[\nu_j(ya_s) - \nu_j(y)]\}^2\, dG_n(y) \end{aligned} \tag{92}$$

where ν_j is as in (8) and

$$\nu_j^*(y) := n^{-1/2} \sum_i d_{nij}\, f_{ni}(y), \qquad 1 \le j \le p. \tag{93}$$

The representation (92) suggesting the following approximating candidate:

$$\hat{K}_D^o(s, u) := \sum_{j=1}^{p} \int \{Y_j^o + u'\nu_j + sy\nu_j^* + m_j\}^2\, dG_n. \tag{94}$$

We now state

Lemma 5.5.5. *With γ_n as in (6), assume that $\forall$ $|s| \le b$, $0 < b < \infty$,*

$$\begin{aligned} \lim_{x \to 0} \limsup_n & \int \gamma_n((1+sn^{-1/2})y+x)\,dG_n(y) \\ & = \limsup_n \int \gamma_n(y)\,dG_n(y) < \infty, \end{aligned} \tag{95}$$

and

$$\text{(96)} \qquad \lim_{z\to 0} \limsup_n \int |y|\, \gamma_n(y+zy)\, dG_n(y)$$
$$= \limsup_n \int |y|\, \gamma_n(y)\, dG_n(y) < \infty.$$

Moreover, assume that $\forall\ (s, \mathbf{v}) \in [-b, b] \times \mathcal{N}(B) =: \mathcal{N}_1$, *and* $\forall\ \delta > 0$

$$\text{(97)} \quad \limsup_n \sum_{j=1}^{p} \int \Big[\sum_{i=1}^{n} d_{nij}^{\pm}\{F_{ni}(ya_s + \mathbf{c}_{ni}'\mathbf{v} + \delta(n^{-1/2}|y| + \kappa_{ni})) -$$
$$- F_{ni}(ya_s + \mathbf{c}_{ni}'\mathbf{v} - \delta(n^{-1/2}|y| + \kappa_{ni}))\}\Big]^2 dG_n(y)$$
$$\le k\delta^2,$$

for some k *not depending on* $(s, \mathbf{v})$ *and* δ.
Then, $\forall\ 0 < b, B < \infty$,

$$\text{(98)} \qquad E \sup \sum_{j=1}^{p} \int \{Y_j^o((1+sn^{-1/2})y, \mathbf{u}) - Y_j^o(y)\}^2\, dG_n = o(1)$$

where the supremum is taken over $(s, \mathbf{u}) \in \mathcal{N}_1$.

Proof. For each $(s, \mathbf{u}) \in \mathcal{N}_1$, with $a_s = 1 + sn^{-1/2}$,

$$E \sum_{j=1}^{p} \int \{Y_j^o(ya_s, \mathbf{u}) - Y_j^o(y)\}^2\, dG_n(y)$$
$$\le \int_{-B_n}^{B_n} \int \gamma_n(ya_s+z)\, dG_n(y)dz + \int_{-b_n}^{b_n} \int |y|\, \gamma_n(y+zy)\, dG_n(y)dz$$

where $B_n = B \max_i \|\kappa_i\|$, $b_n = bn^{-1/2}$. Therefore, from (95) and (96), for every $(s, \mathbf{u}) \in \mathcal{N}_1$,

$$E \sum_{j=1}^{p} \int \{Y_j^o(ya_s, \mathbf{u}) - Y_j^o(y)\}^2\, dG_n(y) = o(1).$$

Now proceed as in the proof of (16), using the monotonicity of $V_{jd}(a, \mathbf{t})$, $\mu_j^o(a, \mathbf{u})$ and the compactness of $\mathcal{N}_1$ to conclude (98). Use (97) in place of (7). The details are left out as an exercise. □

The proof of the following lemma is quite similar to that of (30).

Lemma 5.5.6. *Let* $G_n^*(y) = G_n(y/a_\tau)$. *Assume that for each fixed* $(\tau, \mathbf{u}) \in \mathcal{N}_1$, (8) *and* (9) *hold with* G_n *replaced by* G_n^*. *Moreover, assume the following:*

(99) $$\sum_{j=1}^{p} \int (y\overset{*}{\nu}_j(y))^2 \, dG_n(y) = O(1).$$

(100) $$\sum_{j=1}^{p} \int \{\mu_j^o(ya_s) - \mu_j^o(y) - \tau y \overset{*}{\nu}_j(y)\}^2 \, dG_n(y) = o(1), \quad \forall \ |s| \le b.$$

Then, $\forall\ 0 < b, B < \infty$,

(101) $$\sup \sum_{j=1}^{p} \int \{\mu_j^o(ya_s, \mathbf{u}) - \mu_j^o(a_\tau y) - \mathbf{u}'\nu_j(a_\tau y)\}^2 \, dG_n(y) = o(1),$$

and

(102) $$\sup \sum_{j=1}^{p} \int \{\mu_j^o(ya_s) - \mu_j^o(y) - \tau y \overset{*}{\nu}_j(y)\}^2 \, dG_n(y) = o(1).$$

where the supremum in (101), (102) *is taken over* $(s, \mathbf{u}) \in \mathcal{N}_1$, $|s| \le b$, *respectively.*

Theorem 5.5.8. *Let* $Y_{n1}, \ldots, Y_{nn}$ *be independent r.v.'s with respective d.f.'s* $F_{n1}, \ldots, F_{nn}$. *Assume* (1) – (5), (8), (10), (95) – (97) *and the conditions of Lemma* 5.5.6 *hold. Moreover assume that for each* $|s| \le b$

(103) $$\sum_{j=1}^{p} \int \|\nu_j(ya_s) - \nu_j(y)\|^2 \, dG_n(y) = o(1).$$

Then, $\forall\ 0 < b,\ B < \infty$,

(104) $$E \sup |K_D^o(\tau, \mathbf{u}) - \hat{K}_D^o(\tau, \mathbf{u}) = o(1).$$

where the supremum is taken over $(s, \mathbf{u}) \in \mathcal{N}_1$.

The proof of this theorem is quite similar to that of Theorem 5.5.1. □

5.6. ASYMPTOTIC DISTRIBUTIONS, EFFICIENCES AND ROBUSTNESS

5.6a. Asymptotic Distributions and Efficiences

To begin with consider the *Case* 5.5.1 *and the class of estimators* $\{\hat{\beta}_D\}$. Recall that in this case the errors $\{e_{ni}\}$ of (1.1.1) are correctly modeled to be i.i.d. F, i.e., $H_{ni} \equiv F \equiv F_{ni}$. We shall also take $G_n \equiv G$, $G \in \mathbb{D}\mathcal{I}(\mathbb{R})$. Assume that (5.5.68) – (5.5.70) hold. The various quantities appearing in (5.5.37) and Theorem 5.5.3 now take the following simpler forms.

(1) $$\Gamma_n(y) = A\,X'D\,f(y),\ y \in \mathbb{R}, \qquad \mathscr{B}_n = AX'D\,D'XA \int f\,dG,$$

$$\mathscr{T}_n = -AX'D \int Y^o_D\, f\,dG.$$

Note that $\mathscr{B}_n^{-1}$ will exist if and only if the rank of D is p. Note also that

(2) $$\mathscr{B}_n^{-1}\mathscr{T}_n = -(D'XA)^{-1} \int Y^o_D\, fdG \,/\, (\int f^2dG)^{-1}$$
$$= (D'XA \int f^2dG)^{-1}\, \Sigma_i\, d_i\, [\psi(e_i) - E\psi(e_i)],$$

where $\psi(y) = \int_{-\infty}^{y} fdG, \quad y \in \mathbb{R}$.

Because $G_n \equiv G \in \mathbb{D}\mathcal{I}(\mathbb{R})$, there always exists a $g \in L_r^2(G)$ such that $g > 0$, and $0 < \int g^2dG < \infty$. Take $g_n \equiv g$ in (5.5.11). Then the condition (5.5.11) translates to assuming that

(3) $$\liminf_n \inf_{\|\theta\|=1} |\theta' D'XA\theta| \geq \alpha \qquad \text{for some } \alpha > 0.$$

Condition (5.5.12) implies that $\theta' D'XA\theta \geq 0$ or $\theta' D'XA\,\theta \leq 0$, $\forall\ \|\theta\| = 1$ and $\forall\ n \geq 1$. It need not imply (3). The above discussion together with the L-F Cramer-Wold Theorem leads to

Corollary 5.6a.1. *Assume that* (1.1.1) *holds with the error r.v.'s correctly modeled to be i.i.d.* F, F *known. In addition, assume that* (5.5.1), (5.5.2), (5.5.12), (5.5.68) – (5.5.70), (3) *and* (4) *hold, where*

(4) $$(D'XA)^{-1} \textit{ exists for all } n \geq p.$$

Then,

(5) $$A^{-1}(\hat\beta_D - \beta) = (D'XA \int f^2dG)^{-1} \sum_{i=1}^{n} d_{ni}\, [\psi(e_{ni}) - E\psi(e_{ni})] + o_p(1).$$

If, in addition, we assume

(6) $$\max_{1\leq i\leq n} \|d_{ni}\|^2 = o(1),$$

then

(7) $$\Sigma_D^{-1} A^{-1}(\hat\beta_D - \beta) \xrightarrow[d]{} N(0, \tau^2 I_{p\times p})$$

where

$$\Sigma_D := (D'XA)^{-1} D'D (AX'D)^{-1}, \qquad \tau^2 = \mathrm{Var}\, \psi(e_1)/(\int f^2dG)^2. \qquad \square$$

For any two square matrices L_1 and L_2 of the same order, by $L_1 \geq L_2$ we mean that $L_1 - L_2$ is non–negative definite. Let L and J be two $p{\times}n$ matrices such that $(LL')^{-1}$ exists. The C-S inequality for matrices states that

(8) $$JJ' \geq JL'(LL')^{-1}LJ' \text{ with equality if and only if } J \propto L.$$

Now note that if $D = XA$ then $\Sigma_D = I_{p\times p}$. In general, upon choosing $J = D'$, $L = AX'$ in (8), we obtain

$$D'D \geq D'XA \cdot AX'D \quad \text{or} \quad \Sigma_D \geq I_{p\times p}$$

with equality if and only if $D \propto XA$. From these observations we deduce

Theorem 5.6a.1. (***Optimality of*** $\hat{\beta}_X$). *Suppose that* (1.1.1) *holds with the error r.v.'s correctly modeled to be i.i.d.* F. *In addition, assume that* (5.5.1), (5.5.4) *with* $G_n \equiv G$, (5.5.68) – (5.5.70) *hold. Then, among the class of estimators* $\{\hat{\beta}_D;$ D *satisfying* (5.5.2), (5.5.12), (3), (4) *and* (5)$\}$, *the estimator that minimizes the asymptotic variance of* $b'A^{-1}(\hat{\beta}_D - \beta)$, *for every* $b \in \mathbb{R}^p$, *is* $\hat{\beta}_X$ *– the* $\hat{\beta}_D$ *with* $D = XA$. □

Observe that under (5.5.1), $D = XA$ *a priori* satisfies (5.5.2), (3), (4) and (6). Consequently we obtain

Corollary 5.6a.2. (***Asymptotic normality of*** $\hat{\beta}_X$.) *Assume that* (1.1.1) *holds with the error r.v.'s correctly modeled to be i.i.d.* F. *In addition, assume that* (5.5.1) and (5.5.68) – (5.5.70) *hold. Then,*

$$A^{-1}(\hat{\beta}_X - \beta) \xrightarrow[d]{} N(0, \tau^2 I_{p\times p}). \qquad \square$$

Remark 5.6a.1. Write $\hat{\beta}_D(G)$ for $\hat{\beta}_D$ to emphasize the dependence on G. The above theorem proves the optimality of $\hat{\beta}_X(G)$ among a class of estimators $\{\hat{\beta}_D(G)$, as D varies$\}$. To obtain an asymptotically efficient estimator at a given F among the class of estimators $\{\hat{\beta}_X(G)$, G varies$\}$ one must have F and G satisfy the following relation. Assume that F satisfies (3.2.a) of Theorem 3.2.3 and all of the derivatives that occur below make sense and that (5.5.68) hold. Then, a G that will give asymptotically efficient $\hat{\beta}_X(G)$ must satisfy the relation

$$-f\,dG = (1/I(f))\cdot d(\dot{f}/f), \qquad I(f) := \int (\dot{f}/f)^2 dF.$$

From this it follows that the m.d. estimators $\hat{\boldsymbol{\beta}}_{\mathbf{X}}(G)$, for G satisfying the relations $dG(y) = (2/3)dy$ and $dG(y) = 4d\delta_0(y)$, are asymptotically efficient at logistic and double exponential error d.f.'s, respectively.

For $\hat{\boldsymbol{\beta}}_{\mathbf{X}}(G)$ to be asymptotically efficient at $N(0, 1)$ errors, G would have to satisfy $f(y)dG(y) = dy$. But such a G does not satisfy (5.5.58). Consequently, under the current art of affairs, one can not estimate $\boldsymbol{\beta}$ asymptotically efficiently at the $N(0, 1)$ error d.f. by using a $\hat{\boldsymbol{\beta}}_{\mathbf{X}}(G)$. This naturally leaves one open problem, v.i.z., *Is the conclusion of Corollary* 5.6a.2 *true without requiring* $\int f dG < \infty$, $0 < \int f^2 dG < \infty$? □

Observe that Theorem 5.6a.1 does not include the estimator $\hat{\beta}_1$ – the $\hat{\boldsymbol{\beta}}_{\mathbf{D}}$ when $\mathbf{D} = n^{1/2}[1, 0, \ldots, 0]_{n\times p}$ i.e., the m.d. estimator defined at (5.2.4), (5.2.5) after H_{ni} is replaced by F in there. The main reason for this being that the given $\mathbf{D}$ does not satisfy (4). However, Theorem 5.5.3 is general enough to cover this case also. Upon specializing that theorem and applying (5.5.49) one obtains the following

Theorem 5.6a.2. *Assume that* (1.1.1) *holds with the errors correclty modeled to be i.i.d.* F. *In addition, assume that* (5.5.1), (5.5.68) – (5.5.70) *and the following hold.*

(10) *Either*

$$n^{-1/2}\theta_1 \mathbf{x}_{ni}'\mathbf{A}\boldsymbol{\theta} \geq 0 \;\; \textit{for all} \;\; 1 \leq i \leq n, \;\; \textit{all} \;\; \|\boldsymbol{\theta}\| = 1,$$

or

$$n^{-1/2}\theta_1 \mathbf{x}_{ni}'\mathbf{A}\boldsymbol{\theta} \leq 0 \;\; \textit{for all} \;\; 1 \leq i \leq n, \;\; \textit{all} \;\; \|\boldsymbol{\theta}\| = 1.$$

(11)
$$\liminf_n \inf_{\|\boldsymbol{\theta}\|=1} |n^{1/2}\theta_1 \bar{\mathbf{x}}_n'\mathbf{A}\boldsymbol{\theta}| \geq \alpha > 0,$$

where $\bar{\mathbf{x}}_n$ *is as in* (4.2a.11) *and* θ_1 *is the first coordinate of* $\boldsymbol{\theta}$. *Then*

(12)
$$n^{1/2}\bar{\mathbf{x}}_n'\mathbf{A}\cdot\mathbf{A}^{-1}(\hat{\beta}_1 - \boldsymbol{\beta}) = Z_n \Big/ \int f^2 dG + o_p(1),$$

where

$$Z_n = n^{-1/2}\,\Sigma_i\,\{\psi(e_{ni}) - E\psi(e_{ni})\}, \quad \textit{with } \psi \textit{ as in } (2).$$

Consequently, $n^{1/2}\bar{\mathbf{x}}_n'(\hat{\beta}_1 - \boldsymbol{\beta})$ *is asymptotically a* $N(0, \tau^2)$ *r.v.* □

Next, we focus on the class of estimators $\{\beta_{\mathbf{D}}^{+}\}$ and the *case of i.i.d. symmetric errors.* An analogue of Corollary 5.6a.1 is obtained with the help of Theorem 5.5.4 instead of Theorem 5.5.3 and is given in Corollary 5.6a.3. The details of its proof are similar to those of Corollary 5.6a.1.

Corollary 5.6a.3. *Assume that* (1.1.1) *holds with the errors correctly modeled to be i.i.d. symmetric around* 0. *In addition, assume that* (5.3.8), (5.5.1), (5.5.2), (5.5.4) *with* $G_n \equiv G$, (5.5.68), (5.5.70), (3), (4) *and* (13) *hold, where*

$$(13) \qquad \int_0^{\infty} (1-F)\, dG < \infty$$

Then,

$$(14) \qquad \mathbf{A}^{-1}(\beta_{\mathbf{D}}^{+} - \beta) = -\{2\mathbf{A}\mathbf{X}'\mathbf{D}\int f^2 dG\}^{-1} . \int \mathbf{W}^{+}(y)\, f^{+}(y) dG(y) + o_p(1),$$

where $f^{+}(y) := f(y) + f(-y)$ *and* $\mathbf{W}^{+}(y)$ *is* $\mathbf{W}^{+}(y, 0)$ of (5.5.32). *If, in addition,* (6) *holds, then*

$$(15) \qquad \Sigma_{\mathbf{D}}^{-1}\mathbf{A}^{-1}(\beta_{\mathbf{D}}^{+} - \beta) \xrightarrow[d]{} N(0, \tau^2 I_{p\times p}). \qquad \square$$

Consequently, an analogue of Theorem 5.6a.1 holds for $\beta_{\mathbf{X}}^{+}$ also and Remark 5.6a.1 applies equally to the class of estimators $\{\beta_{\mathbf{X}}^{+}(G),\ G \text{ varies}\}$, assuming that the errors are symmetric around 0. We leave it to interested readers to state and prove an analogue of Theorem 5.6a.2 for β_1^{+}.

Now consider the class of estimators $\{\beta_{\mathbf{D}}^{*}\}$ of (5.2.23). Recall the notation in (5.5.61) and Theorem 5.5.6. The distributions of these estimators will be discussed when the errors in (1.1.1) are correctly modeled to be i.i.d. F, F an arbitrary d.f. and when $L_n \equiv L$. In this case various entities of Theorem 5.5.6 acquire the following forms.

$$\mu_{\mathbf{D}} \equiv 0; \qquad \ell_{ni}(s) \equiv 1; \qquad \mathbf{D}(s) \equiv \mathbf{D}, \text{ under } (5.2.21);$$

$$\Gamma_n^{*}(s) \equiv \mathbf{A}_1\mathbf{X}_c'\mathbf{D}\, q(s), \qquad q = f(F^{-1});$$

$$\mathcal{T}_n^{*} = -\,\mathbf{A}_1\mathbf{X}_c'\mathbf{D}\int \mathbf{Y}_{\mathbf{D}}\, q dL \;=\; \mathbf{A}_1\mathbf{X}_c'\mathbf{D}\sum_{i=1}^{n} \mathbf{d}_{ni}\, \varphi_0(F(e_{ni}));$$

$$\mathcal{B}_n^{*} = (\mathbf{A}_1\mathbf{X}_c'\mathbf{D}\,\mathbf{D}'\mathbf{X}_c\mathbf{A}_1)\int q^2 dL,$$

where $\mathbf{X}_c$ and $\mathbf{A}_1$ are defined at (4.2a.11) and where

$$\varphi_0(u) := \int_0^{u} q(s) dL(s), \qquad 0 \le u \le 1.$$

Arguing as for Corollary 5.6a.1, one obtains the following

Corollary 5.6a.4. *Assume that* (1.1.1) *holds with the errors correctly modeled to be i.i.d.* F *and that* L *is a d.f.. In addition, assume that* (**F1**), ($\mathbf{NX_c}$), (5.2.21), (5.5.2), *and the following hold.*

(16) $$\liminf_n \inf_{\|\boldsymbol{\theta}\|=1} |\boldsymbol{\theta}' \mathbf{D}' \mathbf{X}_c \mathbf{A}_1 \boldsymbol{\theta}| \geq \alpha > 0$$

(17) Either

$$\boldsymbol{\theta}' \mathbf{d}_{ni}(\mathbf{x}_{ni} - \bar{\mathbf{x}}_n)' \mathbf{A}_1 \boldsymbol{\theta} \geq 0, \quad \forall\ 1 \leq i \leq n, \ \forall\ \|\boldsymbol{\theta}\| = 1,$$

or

$$\boldsymbol{\theta}' \mathbf{d}_{ni}(\mathbf{x}_{ni} - \bar{\mathbf{x}}_n)' \mathbf{A}_1 \boldsymbol{\theta} \leq 0, \quad \forall\ 1 \leq i \leq n, \ \forall\ \|\boldsymbol{\theta}\| = 1.$$

(18) $$(\mathbf{D}' \mathbf{X}_c \mathbf{A}_1)^{-1} \text{ exists for all } n \geq p.$$

Then,

(19) $$\mathbf{A}_1^{-1}(\boldsymbol{\beta}_{\mathbf{D}}^* - \boldsymbol{\beta}) = (\mathbf{D}' \mathbf{X}_c \mathbf{A}_1 \int_0^1 q^2 dL)^{-1} \sum_{n=1}^{n} \mathbf{d}_{ni}\, \varphi(F(e_{ni})) + o_p(1).$$

If, in addition, (6) *holds, then*

(20) $$(\boldsymbol{\Sigma}_{\mathbf{D}}^*)^{-1} \mathbf{A}_1^{-1}(\boldsymbol{\beta}_{\mathbf{D}}^* - \boldsymbol{\beta}) \underset{d}{\longrightarrow} N(\mathbf{0}, \sigma_0^2 \mathbf{I}_{p \times p})$$

where $\boldsymbol{\Sigma}_{\mathbf{D}}^* = (\mathbf{D}' \mathbf{X}_c \mathbf{A}_1)^{-1} \mathbf{D}' \mathbf{D} (\mathbf{A}_1 \mathbf{X}_c' \mathbf{D})^{-1}, \quad \sigma_0^2 = \operatorname{Var} \varphi(F(e_1)) / (\int_0^1 q^2 dL)^2.$
Consequently,

(21) $$\mathbf{A}_1^{-1}(\boldsymbol{\beta}_{\mathbf{X}_c}^* - \boldsymbol{\beta}) \underset{d}{\longrightarrow} N(\mathbf{0}, \sigma_0^2 \mathbf{I}_{p \times p})$$

and $\{\boldsymbol{\beta}_{\mathbf{X}_c}^*\}$ *is asymptotically efficient among all* $\{\boldsymbol{\beta}_{\mathbf{D}}^*$, **D** *satisfying above conditions*$\}$. □

Consider the case when $L(s) \equiv s$. Then

$$\sigma_0^2 = \left(\int f^3(x)dx\right)^{-2} \int\int [F(x \wedge y) - F(x)F(y)] f^2(x) f^2(y)\, dx dy.$$

It is interesting to make a numerical comparison of this variance with that of some other well celebrated estimators. Let σ_w^2, σ_{lad}^2, σ_{ls}^2 and σ_{ns}^2 denote the respective asymptotic variances of the Wilcoxon rank, the least absolute deviation, the least square and the normal scores estimators of $\boldsymbol{\beta}$. Recall, either from Chapter 4 or from Jaeckel (1972) that

$$\sigma_{\mathrm{w}}^2 = (1/12)\cdot\{\int f^2(x)\,dx\}^{-2};\quad \sigma_{\mathrm{lad}}^2 = (2\,f(0))^{-2};\quad \sigma_{\mathrm{ls}}^2 = \sigma^2;$$

$$\sigma_{\mathrm{ns}}^2 = \{\int f^2(x)/\varphi(\Phi^{-1}(F))\,dx\}^{-2};$$

where σ^2 is the error variance. Using these we obtain the following table.

Table I

F \ $\sigma_{\cdot}^2$	σ_0^2	σ_{w}^2	σ_{lad}^2	σ_{ns}^2	σ^2
Double Exp.	1.2	1.333	1	$\pi/2$	2
Logistic	3.0357	3	4	π	$\pi^2/3$
Normal	1.0946	$\pi/3$	$\pi/2$	1	1

It thus follows that the m.d. estimator $\beta^*_{\mathbf{X}_c}(L)$, with $L(s) \equiv s$, is superior to the Wilcoxon rank estimator and the l.a.d. estimator at double exponential and logistic errors, respectively. At normal errors, it has smaller variance than the l.a.d. estimator and compares favorably with the optimal estimator. The same is true for the m.d. estimator $\hat{\beta}_{\mathbf{X}}(F)$.

Next, we shall discuss $\bar{\beta}$. In the following theorem the framework is the same as in Theorem 5.5.7. Also see (5.5.82) for the definitions of $\bar{\nu}_n$, $\mathscr{B}_{\mathrm{ln}}$ etc.

Theorem 5.6a.3. *In addition to the assumptions of Theorem* 5.5.7 *assume that*

(22) $$\liminf_n \inf_{\|\theta\|=1} |\int \bar{\nu}_n'\, d\alpha_n \theta| \geq \alpha, \qquad \textit{for some } \alpha > 0.$$

Moreover, assume that (10) *holds and that*

(23) $\mathscr{B}_{\mathrm{ln}}^{-1}$ *exists for all* $n \geq p$.

Then,

(24) $$A^{-1}(\bar{\beta} - \beta) = -\mathscr{B}_{\mathrm{ln}}^{-1}\int\int \bar{\nu}_n(s, y)\{\mathscr{M}_{\mathrm{ln}}(s, y) + m_n(s, y)\}\, d\alpha_n(s, y) + o_p(1).$$

Proof. The proof of (23) is similar to that of (5.5.51), hence no details are given. □

Corollary 5.6a.5. *Suppose that the conditions of Theorem* 5.6a.3 *are satisfied by* $F_{ni} \equiv F \equiv H_{ni}$, $G_n \equiv G$, $L_n \equiv L$, *where* F *is supposed to have*

continuous density f. *Let*

$$(25)\quad C=\int\int\int_0^1\int_0^1[\{A\, n^{-1}\sum_{i=1}^{ns}\sum_{j=1}^{nt} \mathbf{x}_i\mathbf{x}_j' f_i(y)f_j(y)\, A\}\,(s\wedge t)\cdot$$
$$\cdot\{F(y\wedge z)-F(y)F(z)\}]\, d\alpha(s,y)\, d\alpha(t,z),$$

where $f_i(y) = f(y-\mathbf{x}_i'\beta)$, *and* $d\alpha(s, y) = dL(s)dG(y)$. *Then the asymptotic distribution of* $A^{-1}(\bar\beta-\beta)$ is $N(0, \Sigma_0(\beta))$ *where* $\Sigma_0(\beta) = \mathscr{B}_{1n}^{-1}C\mathscr{B}_{1n}^{-1}$. □

Because of the dependence of Σ_0 on β, no clear cut comparison between $\bar\beta$ and $\hat\beta_{\mathbf{X}}$ in terms of their asymptotic covariance matrices seems to be feasible. However, some comparison at a given β can be made. To demonstrate this, consider the case when $L(s) = s$, $p = 1$ and $\beta_1 = 0$. Write x_i for x_{i1} etc.

Note that here, with $\tau_x^2 = \sum_{i=1}^{n} x_i^2$,

$$\mathscr{B}_{1n} = \tau_x^{-2}\int_0^1 n^{-1}\sum_{i=1}^{ns} x_i\sum_{j=1}^{ns} x_j\, ds\cdot\int f^2\, dG,$$

$$C=\tau_x^{-2}\int_0^1\int_0^1 n^{-1}\sum_{i=1}^{ns} x_i\sum_{j=1}^{nt} x_j\,(s\wedge t)\, dsdt\cdot$$
$$\cdot\int\int[F(y\wedge z)-F(y)F(z)]\, d\psi(y)\, d\psi(z).$$

Consequently

$$\Sigma_0(0) = \frac{\tau_x^{-2}\int_0^1\int_0^1 (s\wedge t)n^{-1}\sum_{i=1}^{ns} x_i\sum_{j=1}^{nt} x_j\, dsdt}{(\tau_x^{-2}\int_0^1 n^{-1}\sum_{i=1}^{ns} x_i\sum_{j=1}^{ns} x_j\, ds)^2}.\tau^2 = r_n\cdot\tau^2, \qquad \text{say.}$$

Recall that τ^2 is the asymptotic variance of $\tau_x(\hat\beta_x-\beta)$. Direct integration shows that in the cases $x_i\equiv 1$ and $x_i\equiv i$, $r_n \longrightarrow 18/15$ and $50/21$, respectively. Thus, in the cases of the one sample location model and the first degree polynomial through the origin, in terms of the asymptotic variance, $\hat\beta_x$ dominates $\bar\beta$ with $L(s) = s$ at $\beta = 0$. □

5.6b. Robustness

In a linear regression setup an estimator needs to be robust against departures in the assumed design variables and the error distributions. As seen in Section 5.6a, one purpose of having general weights $\mathbf{D}$ in $\hat{\beta}_{\mathbf{D}}$ was to prove that $\hat{\beta}_{\mathbf{X}}$ is asymptotically efficient among a certain class of m.d. estimators $\{\hat{\beta}_{\mathbf{D}}, \mathbf{D}$ varies$\}$. Another purpose is to robustify these estimators against the extremes in the design by choosing $\mathbf{D}$ to be a bounded function of $\mathbf{X}$ that satisfies all other conditions of Theorem 5.6a.1. Then the corresponding $\hat{\beta}_{\mathbf{D}}$ would be asymptotically normal and robust against the extremes in the design, but not as efficient as $\hat{\beta}_{\mathbf{X}}$. This gives another example of the phenomenon that compromises efficency in return for robustness. A similar remark applies to $\{\beta_{\mathbf{D}}^{+}\}$ and $\{\beta_{\mathbf{D}}^{*}\}$.

We shall now focus on the *qualitative robustness* (see Definition 4.4.1) of $\hat{\beta}_{\mathbf{X}}$ and $\beta_{\mathbf{X}}^{+}$. For simplicity, we shall write $\hat{\beta}$, β^{+}, for $\hat{\beta}_{\mathbf{X}}$, $\beta_{\mathbf{X}}^{+}$ in the rest of the section. To begin with consider $\hat{\beta}$. Recall Theorem 5.5.3 and the notation of (5.5.37). We need to apply these to the case when the errors in (1.1.1) are modeled to be i.i.d. F, but their actual d.f.'s are $\{F_{ni}\}$, $\mathbf{D} = \mathbf{XA}$ and $G_n = G$. Then various quantities in (5.5.37) acquire the following form.

$$(1) \qquad \Gamma_n(y) = \mathbf{AX}'\Lambda^*(y)\mathbf{XA}, \qquad \mathscr{B}_n = \mathbf{AX}'\int \Lambda^*\Pi\Lambda^*\, dG\, \mathbf{XA},$$

$$\mathscr{T}_n = \int \Gamma_n(y)\mathbf{AX}'[\boldsymbol{\alpha}_n(y) + \Delta_n(y)]\, dG(y) = \mathbf{Z}_n + \mathbf{b}_n, \qquad \text{say,}$$

where

$$(2) \qquad \Pi := \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'; \qquad \mathbf{b}_n := \int \Gamma_n(y)\mathbf{AX}'\Delta_n(y)\, dG(y);$$

$$\alpha_{ni}(y) := I(e_{ni} \le y) - F_{ni}(y),$$

$$\Delta_{ni}(y) := F_{ni}(y) - F(y), \qquad 1 \le i \le n, \ y \in \mathbb{R};$$

$$\boldsymbol{\alpha}_n' := (\alpha_{n1}, \alpha_{n2}, ..., \alpha_{nn}), \qquad \Delta_n' := (\Delta_{n1}, \Delta_{n2}, ..., \Delta_{nn}).$$

The assumption (5.2.1) ensures that the design matrix $\mathbf{X}$ is of the full rank p. This in turn implies the existence of $\mathscr{B}_n^{-1}$ and the satisfaction of (5.2.2), (5.2.12) in the present case. Moreover, because $G_n \equiv G$, (5.2.11) now becomes

$$(3) \qquad \liminf_n \inf_{\|\theta\|=1} k_n(\theta) \geq \gamma, \qquad \text{for some } \gamma > 0,$$

where

$$k_n(\theta) := \theta' AX' \int \Lambda^* \, g dG \, XA\theta, \qquad \|\theta\| = 1,$$

and where g is a function from $\mathbb{R}$ to $[0, \infty]$, $0 < \int g^r \, dG < \infty$, $r = 1, 2$. Because G is a σ–finite measure, such a g always exists.

Upon specializing Theorem 5.5.3 to the present case, we readily obtain

Corollary 5.6b.1. *Assume that* in (1.1.1) *the actual and modeled d.f.'s of the errors* $\{e_{ni}, 1 \leq i \leq n\}$ *are* $\{F_{ni}, 1 \leq i \leq n\}$ *and* F, *respectively. In addition, assume that* (5.5.1), (5.5.3) – (5.5.10) *with* $D = XA$, $H_{ni} \equiv F$, $G_n \equiv G$, *and* (3) *hold. Then*

$$(4) \qquad A^{-1}(\hat{\beta} - \beta) = -\mathscr{B}_n^{-1}\{Z_n + b_n\} + o_p(1). \qquad \square$$

Observe that $\mathscr{B}_n^{-1} b_n$ measures the amount of the asymptotic bias in the estimator $\hat{\beta}$ when $F_{ni} \neq F$. Our goal here is to obtain the asymptotic distribution of $A^{-1}(\hat{\beta} - \beta)$ when $\{F_{ni}\}$ converge to F in a certain sense. The achievement of this goal is facilitated by the following lemma. Recall that for any square matirx L, $\|L\|_\infty = \sup\{\|t'L\|; \|t\| \leq 1\}$. Also recall the fact that

$$(5) \qquad \|L\|_\infty \leq \{\mathrm{tr}.LL'\}^{1/2},$$

where tr. denotes the trace operator.

Lemma 5.6b.1. *Let* F *and* G *satisfy* (5.5.68). *Assume that* (5.5.5) *and* (5.5.10) *are satisfied by* $G_n \equiv G$, $\{F_{ni}\}$, $H_{ni} \equiv F$ *and* $D = XA$. *Moreover assume that* (5.5.3) *holds and that*

$$(6) \qquad \rho_n := \int (\Sigma_i \|x_{ni}'A\|^2 \, |f_{ni} - f|)^2 dG = o(1).$$

Then with $I = I_{p \times p}$,

$$(i) \qquad \|\mathscr{B}_n - I\int f^2 dG\|_\infty = o(1).$$

$$(ii) \qquad \|\mathscr{B}_n^{-1} - I(\int f^2 dG)^{-1}\|_\infty = o(1).$$

$$(iii) \qquad |\mathrm{tr}.\ \mathscr{B}_n - p\int f^2 dG| = o(1).$$

(iv) $$|\sum_{j=1}^{p} \int \|\nu_j\|^2 dG - p\int f^2 dG| = o(1).$$

(v) $$\|b_n - \int AX' \Delta_n(y)f(y)dG(y)\| = o(1).$$

(vi) $$\|Z_n - \int AX' \alpha_n(y)f(y)dG(y)\| = o_p(1).$$

(vii) $$\sup_{\|\theta\|=1} |k_n(\theta) - \int f\, g dG| = o(1).$$

Remark 5.6b.1. Note that the condition (5.5.10) with $D = XA$, $G_n \equiv G$ now becomes

(7) $$\int \|AX' \Delta_n\|^2 dG = O(1).$$

Proof. To begin with, because $AX'XA \equiv I$, we obtain the relation

$$\begin{aligned}\Gamma_n(y)\Gamma_n'(y) - f^2(y)I &= AX'[\Lambda^*(y) - f(y)I]XA \cdot AX'[\Lambda^*(y) - f(y)I]XA \\ &= AX'\mathcal{C}(y)XA \cdot AX'\mathcal{C}(y)XA \\ &= \mathcal{D}(y)\mathcal{D}'(y), \qquad y \in \mathbb{R},\end{aligned}$$

where $\mathcal{C}(y) := \Lambda^*(y) - I\, f(y)$, $\mathcal{D}(y) := AX'\mathcal{C}(y)XA$, $y \in \mathbb{R}$. Therefore,

(8) $$\| \mathcal{B}_n - I\int f^2 dG\|_\infty \le \sup_{\|t\|\le 1} \int \|t'\mathcal{D}(y)\mathcal{D}'(y)\|\, dG(y) \le \int \{tr.LL'\}^{1/2} dG$$

where $L = \mathcal{D}\mathcal{D}'$. Note that, by the C–S inequality,

(9) $$tr.\ LL' = tr.\ \mathcal{D}\mathcal{D}'\mathcal{D}'\mathcal{D} \le \{tr.\mathcal{D}\mathcal{D}'\}^2.$$

Let $\delta_i = f_i - f$, $1 \le i \le n$. Then

(10) $$\begin{aligned}|tr.\mathcal{D}\mathcal{D}'| &= |tr.\ \Sigma_i \Sigma_j\ Ax_i\, x_i'A \cdot Ax_j x_j'A \cdot \delta_i\delta_j| \\ &= |\Sigma_i \Sigma_j\ \delta_i\delta_j\ (x_j'AAx_i)^2| \\ &\le \Sigma_i \Sigma_j\ |\delta_i\delta_j| \cdot \|x_i'A\|^2 \cdot \|x_j'A\|^2 \\ &= (\Sigma_i \|Ax_i\|^2 |\delta_i|)^2 = \rho_n.\end{aligned}$$

Consequently, from (8) – (10),

$$\|\mathcal{B}_n - \mathbf{I}\int f^2 dG\|_\infty \le \int (\Sigma_i \|\mathbf{A}\mathbf{x}_i\|^2 |f_i - f|)^2 dG = o(1), \quad \text{by (6).} \tag{11}$$

This proves (i) while (ii) follows from (i) by using the determinant and cofactor formula for the inverses.

Next, (iii) follows from (6) and the fact that

$$|\text{tr.}\mathcal{B}_n - p\int f^2 dG| = |\int \text{tr.}\mathcal{D}\mathcal{D}'\, dG| \le \rho_n, \qquad \text{by (10).} \tag{12}$$

To prove (iv), note that with $\mathbf{D} = \mathbf{X}\mathbf{A}$,

$$\sum_{j=1}^{p} \int \|\nu_j\|^2 dG = \sum_{i=1}^{n} \sum_{k=1}^{n} \int \mathbf{x}_i'\mathbf{A}\mathbf{A}\mathbf{x}_k\, \mathbf{x}_k'\mathbf{A}\mathbf{A}\mathbf{x}_i\, f_i(y) f_k(y)\, dG(y).$$

Note that the r.h.s. is $p\int f^2 dG$ in the case $f_i \equiv f$. Thus

$$|\sum_{j=1}^{p} \int \|\nu_j\|^2\, dG - p\int f^2 dG| = |\int \text{tr.}\mathcal{D}\mathcal{D}' dG| \le \rho_n. \tag{13}$$

This and (6) proves (iv).

Similarly, with $d_j'(y)$ denoting the j^{th} row of $\mathcal{D}(y)$, $1 \le j \le p$,

$$\begin{aligned} \|\mathbf{b}_n - \int \mathbf{A}\mathbf{X}'\Delta_n\, f dG\|^2 &= \|\int \mathcal{D}\mathbf{A}\mathbf{X}'\Delta\, dG\|^2 \\ &= \sum_{j=1}^{p} \{\int d_j'(y)\mathbf{A}\mathbf{X}'\Delta_n(y)\, dG(y)\}^2 \\ &\le \rho_n \int \|\mathbf{A}\mathbf{X}'\Delta_n(y)\|^2\, dG(y) \end{aligned} \tag{14}$$

and

$$\|\mathbf{Z}_n - \int \mathbf{A}\mathbf{X}'\alpha_n(y)\, f(y) dG(y)\|^2 \le \rho_n \int \|\mathbf{A}\mathbf{X}'\alpha_n\|^2 dG. \tag{15}$$

Moreover,

$$E\int \|\mathbf{A}\mathbf{X}'\alpha_n\|^2\, dG = \int \Sigma_i \|\mathbf{x}_i'\mathbf{A}\|^2\, F_i(1 - F_i)\, dG. \tag{16}$$

Consequently, (v) follows from (6), (7) and (14) whereas (vi) follows from (5.5.5), (6), (15) and (16). Finally, with $\mathcal{D}^{1/2} = \mathbf{A}\mathbf{X}'\mathcal{C}^{1/2}$, $\forall\ \boldsymbol{\theta}$,

$$|\mathrm{k_n}(\boldsymbol{\theta}) - \int \mathrm{f\,gdG}| = |\boldsymbol{\theta}' \int \boldsymbol{\mathcal{D}}\,\mathrm{gdG}\,\boldsymbol{\theta}| = \int \|\boldsymbol{\theta}'\boldsymbol{\mathcal{D}}^{1/2}\|^2\,\mathrm{gdG}.$$

Therefore,

$$\sup_{\|\boldsymbol{\theta}\|=1} |\mathrm{k_n}(\boldsymbol{\theta}) - \int \mathrm{f\,gdG}| \le \int \{\Sigma_i \|\mathbf{Ax}_i\|^2 |\mathrm{f_i} - \mathrm{f}|\}\,\mathrm{gdG}$$

$$\le \rho_n \{\int \mathrm{g}^2\,\mathrm{dG}\}^{1/2} = o(1), \qquad \text{by (6).} \qquad \square$$

Corollary 5.6b.2. *Assume that* (1.1.1) *holds with the actual and the modeled d.f.'s of* $\{e_{ni},\ 1 \le i \le n\}$ *equal to* $\{F_{ni},\ 1 \le i \le n\}$ *and* F, *respectively. In addition, assume that* (5.5.1), (5.5.3) – (5.5.7), (5.5.9), (5.5.10) *with* $\mathbf{D} = \mathbf{XA}$, $H_{ni} \equiv F$, $G_n \equiv G$; (5.5.68) *and* (6) *hold.*

Then, (5.5.8) *and* (2) *are satisfied and*

$$\mathbf{A}^{-1}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = -\left(\int f^2 dG\right)^{-1} \{\hat{\mathbf{Z}}_n + \hat{\mathbf{b}}_n\} + o_p(1) \tag{17}$$

where

$$\hat{\mathbf{Z}}_n := \int \mathbf{AX}'\boldsymbol{\alpha}_n(y)\,d\psi(y) = \mathbf{A}\,\Sigma_i\,\mathbf{x}_{ni}\,[\psi(e_{ni}) - \int \psi(x)dF_{ni}(x)],$$

$$\hat{\mathbf{b}}_n := \int \mathbf{AX}'\Delta_n(y)\,d\psi(y) = \int \Sigma_i\,\mathbf{Ax}_{ni}\,[F_{ni} - F]\,d\psi,$$

with ψ *as in* (5.6a.2). $\square$

Consider $\hat{Z}_n$. Note that with $\sigma_{ni}^2 := \mathrm{Var}\{\psi(e_{ni})|F_{ni}\}$, $1 \le i \le n$,

$$\mathrm{E}\,\hat{\mathbf{Z}}_n\hat{\mathbf{Z}}_n' = \Sigma_i\,\mathbf{Ax}_{ni}\mathbf{x}_{ni}'\mathbf{A}\cdot\sigma_{ni}^2.$$

One can rewrite

$$\sigma_{ni}^2 = \int\int [F_{ni}(x \wedge y) - F_{ni}(x)F_{ni}(y)]\,d\psi(x)d\psi(y), \qquad 1 \le i \le n.$$

By (5.5.68a), ψ is nondecreasing and bounded. Hence $\max_i \|F_{ni} - F\|_\infty \longrightarrow 0$ readily implies that $\max_i \sigma_{ni}^2 \longrightarrow \sigma^2$, $\sigma^2 := \mathrm{Var}\{\psi(e)|F\}$. Moreover, we have the inequality

$$|\mathrm{E}\hat{\mathbf{Z}}_n\hat{\mathbf{Z}}_n' - \sigma^2\mathbf{I}_{p\times p}| \le \sum_i \|\mathbf{Ax}_{ni}\|^2\,|\sigma_{ni}^2 - \sigma^2|.$$

It thus readily follows from the L–F CLT that (5.5.1) implies that $\hat{\mathbf{Z}}_n \xrightarrow{d} N(0, \sigma^2\mathbf{I}_{p\times p})$, if $\max_i \|F_{ni} - F\|_\infty \longrightarrow 0$. Consequently, we have

Theorem 5.6b.1. (*Qualitative Robustness*). *Assume the same setup and conditions as in Corollary* 5.6b.2. *In addition, suppose that*

$$\max_i \|F_{ni} - F\|_\infty = o(1), \tag{18}$$

$$\|A\|_\infty = o(1). \tag{19}$$

Then, the distribution of $\hat{\beta}$ *under* $\prod_{i=1}^{n} F_{ni}$ *converges weakly to the degenerate distribution, degenerate at* β.

Proof. It suffices to show that the asymptotic bias is bounded. To that effect we have the inequality

$$\|(\int f^2 dG)^{-1}\, \hat{b}_n\|^2 \le \int \|AX'\Delta\|^2\, dG < \infty, \qquad \text{by (7).}$$

From this, (17), and the above discussion about $\{\hat{Z}_n\}$, we obtain that $\forall\ \eta > 0\ \exists\ K_\eta$ such that $P^n(E_\eta) \longrightarrow 1$, where P^n denotes the probability under $\prod_{i=1}^{n} F_{ni}$ and $E_\eta = \{\|A^{-1}(\hat{\beta} - \beta)\| \le K_\eta\}$. Theorem now follows from this and the elementary inequality $\|\hat{\beta} - \beta\| \le \|A\|_\infty \|A^{-1}(\hat{\beta} - \beta)\|$. □

Remark 5.6b.2. The conditions (6) and (18) together need not imply (5.5.7), (5.5.9) and (5.5.10). The condition (5.5.10) is heavily dependent on the rate of convergence in (18). Note that

$$\|\hat{b}_n\|^2 \le \min\{\psi(\infty)\int \|AX'\Delta\|^2 d\psi,\ (\int f^2 dG)\int \|AX'\Delta\|^2 dG\}. \tag{20}$$

This inequality shows that because of (5.5.68), it is possible to have $\|\hat{b}_n\|^2 = O(1)$ even if (7) (or (5.5.10) with $D = XA$) may not be satisfied. However, our general theory requires (7) any way.

Now, with $\varphi = \psi$ or G,

$$\begin{aligned} \int \|AX'\Delta\|^2 d\varphi &= \int \Sigma_i \Sigma_j\, x_i' A A x_j\, \Delta_i\, \Delta_j\, d\varphi \\ &\le \int (\Sigma_i \|Ax_i\|\, |\Delta_i|)^2 d\varphi. \end{aligned} \tag{21}$$

Thus, if

$$\Sigma_i \|Ax_i\|\ |F_i(y) - F(y)| \le k\, \Delta_n^*(y), \qquad y \in \mathbb{R}, \tag{22}$$

where k is a constant and Δ_n^* is a function such that

$$\text{(23)} \qquad \limsup_n \int (\Delta_n^*)^2 d\varphi < \infty,$$

then (7) would be satisfied and in view of (20), $\|\hat{b}_n\| = O(1)$.

Inequality (22) clearly shows that not every sequence $\{F_{ni}\}$ satisfying (6), (18) and (5.5.3) – (5.5.9) with $D = XA$ will satisfy (7). The rate at which $F_{ni} \Rightarrow F$ is crucial for the validity of (7) or (22). □

We now discuss two interesting examples.

Example 5.6b.1. $F_{ni} = (1 - \delta_{ni})F + \delta_{ni} F_0$, $1 \le i \le n$. This is the Case 5.5.2. From the Claim 5.5.3, (5.5.5) – (5.5.9) are satisfied by this model as long as (5.5.68) – (5.5.70) and (5.5.1) hold. To see if (6) and (7) are satisfied, note that here

$$\rho_n = \int (\Sigma_i \|\mathbf{Ax}_i\|^2 \delta_i |f - f_0\|)^2 dG \le 2 \max_i \delta_i^2 p^2 \cdot [\int (f^2 + f_0^2) dG],$$

and

$$\Sigma_i \|\mathbf{Ax}_i\| \, |F_i - F| = \Sigma_i \|\mathbf{Ax}_i\| \delta_i \, |F - F_0|.$$

Consequently, here (6) is implied by (5.5.68) for (f, G), (f_0, G) and by (5.5.71), while (7) follows from (5.5.72), (21)–(23) upon taking $\Delta_n^* \equiv |F - F_0|$, provided we additionally assume that

$$\text{(24)} \qquad \Sigma_i \|\mathbf{Ax}_i\| \delta_i = O(1).$$

There are two obvious choices of $\{\delta_i\}$ that satisfy (24). They are:

$$\text{(25)} \qquad \text{(a)} \quad \delta_{ni} = n^{-1/2} \quad \text{or} \quad \text{(b)} \quad \delta_{ni} = p^{-1/2} \|\mathbf{Ax}_{ni}\|, \qquad 1 \le i \le n.$$

The gross error models with $\{\delta_i\}$ given by (25b) are more natural than those given by (25a) to linear regression models with unbounded designs. We suggest that in these models, a proportion of contamination one can allow for the i^{th} observation is $p^{-1/2}\|\mathbf{Ax}_i\|$. If δ_i is larger than this in the sense that $\Sigma_i \|\mathbf{Ax}_i\| \delta_i \longrightarrow \infty$ then the bias of $\hat{\beta}$ blows up.

Note that if G is a finite measure, f uniformly continuous and $\{\delta_i\}$ are given by (25b) then all the conditions of the above theorem are satisfied by the above $\{F_i\}$ and F. Thus we have

Corollary 5.6b.3. *Every* $\hat{\beta}$ *corresponding to a finite measure* G *is qualitatively robust for* β *against hetroscadastic gross errors at all those* F's *which have uniformly continuous densities provided* $\{\delta_i\}$ *are given by* (25b) *and provided* (5.5.1) *and* (19) *hold.* □

Example 5.6b.2. Here we consider $\{F_{ni}\}$ given in the Case 5.5.3. We leave it to the reader to verify that one choice of $\{\sigma_{ni}\}$ that implies (7) is to take

$$\sigma_{ni} = \|\mathbf{A}\mathbf{x}_{ni}\|, \qquad 1 \leq i \leq n. \tag{26}$$

One can also verify that in this case, (5.5.68) – (5.5.70), (5.5.75) and (5.5.76) entail the satisfaction of all the conditions of Theorem 5.6b.1. Again, the following corollary holds.

Corollary 5.6b.4. *Every* $\hat{\boldsymbol{\beta}}$ *corresponding to a finite measure* G *is qualitatively robust for* $\boldsymbol{\beta}$ *against hetroscedastic scale errors at all those* F's *which have uniformly continuous densities provided* $\{\sigma_{ni}\}$ *are given by* (26) *and provided* (5.5.1) *and* (19) *hold.* □

As an example of a σ–finite G with $G(\mathbb{R}) = \infty$ that yields a robust estimator, consider $G(y) \equiv (2/3)y$. Assume that the following hold.

(i) F, F_0 have continuous densities f, f_0; $0 < \int f^2 d\lambda, \int f_0^2\, d\lambda < \infty$.

(ii) $\int F(1-F)\, d\lambda < \infty$. (iii) $\int |F - F_0|\, d\lambda < \infty$.

Then the corresponding $\hat{\boldsymbol{\beta}}$ is qualitatively robust at F against the heteroscedastic gross errors of Example 5.6b.1 with $\{\delta_{ni}\}$ given by (25b).

Recall, from Remark 5.6a.1, that this $\hat{\boldsymbol{\beta}}$ is also asymptotically efficient at logistic errors. Thus we have a m.d. estimator $\hat{\boldsymbol{\beta}}$ that is asymptotically efficient *and* qualitatively robust at logistic error d.f. against the above gross errors models!!

We leave it to an interested reader to obtain analogues of the above results for $\boldsymbol{\beta}^+$ and $\boldsymbol{\beta}^*$. The reader will find Theorems 5.5.4 and 5.5.6 useful here. □

5.6c Locally Asymptotically Minimax Property

In this subsection we shall show that the class of m.d. estimators $\{\boldsymbol{\beta}^+\}$ are locally asymptotically minimax (l.a.m.) in the Hájek – Le Cam sense (Hájek (1972), Le Cam (1972)). In order to achieve this goal we need to recall an inequality from Beran (1982) that gives a lower bound on the local asymptotic minimax risk for estimators of Hellinger differentiable functionals on the class of product probability measures. Accordingly, let Q_{ni}, P_{ni} be probability measures on $(\mathbb{R}, \mathscr{B})$, μ_{ni}, ν_{ni} be a σ–finite measures on $(\mathbb{R}, \mathscr{B})$ with ν_{ni} dominating Q_{ni}, P_{ni}; $q_{ni} := dQ_{ni}/d\nu_{ni}$, $p_{ni} := dP_{ni}/d\nu_{ni}$; $1 \leq i \leq n$. Let $Q^n = Q_{n1} \times \dots \times Q_{nn}$ and $P^n = P_{n1} \times \dots \times P_{nn}$ and Π^n denote the class of all n–fold product probability measures $\{Q^n\}$ on $(\mathbb{R}^n, \mathscr{B}^n)$.

Define, for a $c > 0$ and for sequences $0 < \eta_{n1} \longrightarrow 0$, $0 < \eta_{n2} \longrightarrow 0$,

$$\mathscr{H}_n(P^n,c) = \{Q^n \in \Pi^n;\ \Sigma_i \int (q_{ni}^{1/2} - p_{ni}^{1/2})^2 d\nu_{ni} \le c^2\},$$

$$\mathscr{K}_n(P^n,c,\eta_n) = \{Q^n \in \Pi^n;\ Q^n \in \mathscr{H}_n(P^n,c),\ \max_i \int (q_{ni} - p_{ni})^2 d\mu_{ni} \le \eta_{n1},$$
$$\max_i \int (q_{ni}^{1/2} - p_{ni}^{1/2})^2 d\nu_{ni} \le \eta_{n2}\},$$

where $\eta_n' := (\eta_{n1}, \eta_{n2})$.

DEFINITION 5.6c.1. A sequence of vector valued functionals $\{S_n: \Pi^n \to \mathbb{R}^p, n \ge 1\}$ is Hellinger–(H–) differentiable at $\{P^n \in \Pi^n\}$ if there exists a triangular array of $p\times 1$ random vectors $\{\xi_{ni}, 1 \le i \le n\}$ and a sequence of $p\times p$ matrices $\{A_n, n \ge 1\}$ having the following properties:

(i) $\int \xi_{ni} dP_{ni} = 0$, $\int \|\xi_{ni}\|^2 dP_{ni} < \infty$, $1 \le i \le n$; $\quad \Sigma_i \int \xi_{ni}\xi_{ni}' \, dP_{ni} \equiv I_{p\times p}$.

(ii) For every $0 < c < \infty$, every sequence $\eta_n \longrightarrow 0$,

$$\sup \| A_n\{S_n(Q^n) - S_n(P^n)\} - 2\, \Sigma_i \int \xi_{ni}\, p_{ni}^{1/2}(q_{ni}^{1/2} - p_{ni}^{1/2})\, d\nu_{ni} \| = o(1)$$

where the supremum is over all $Q^n \in \mathscr{K}_n(P^n,c,\eta_n)$.

(iii) For every $\epsilon > 0$ and every $\alpha \in \mathbb{R}^p$, with $\|\alpha\| = 1$,

$$\Sigma_i \int (\alpha' \xi_{ni})^2 I(|\alpha' \xi_{ni}| > \epsilon)\, dP_{ni} = o(1).$$

Now, let $X_{n1}, \ldots, X_{nn}$ be independent r.v.'s with $Q_{n1}, \ldots, Q_{nn}$ denoting their respective distributions and $\hat{S}_n = \hat{S}_n(X_{n1}, \ldots, X_{nn})$ be an estimator of $S_n(Q^n)$. Let $\mathscr{U}$ be a nondecreasing bounded function on $[0, \infty]$ to $[0, \infty)$ and define the risk of estimating S_n by $\hat{S}_n$ to be

(1) $$R_n(\hat{S}_n, Q^n) = E^n\{\, \mathscr{U}(\| A_n\{\hat{S}_n - S_n(Q^n)\}\|\},$$

where E^n is the expectation under Q^n.

Theorem 5.6c.1. *Suppose that* $\{S_n: \Pi^n \to \mathbb{R}^p, n \ge 1\}$ *is a sequence of H–differentiable functionals and that the sequence* $\{P^n \in \Pi^n\}$ *is such that*

(2) $$\max_i \int p_{ni}^2 \, d\mu_{ni} = O(1).$$

Then,

(3) $$\lim_{c\to 0} \liminf_n \inf_{\hat{S}_n} \sup_{Q^n \in \mathscr{K}_n(P^n,c,\eta_n)} R_n(\hat{S}_n, Q^n) \geq E\, \mathscr{U}(\|Z\|)$$

where Z *is a* $N(0, I_{pxp})$ *r. v.*

Sketch of a proof. This is a reformulation of a result of Beran (1982), pp 425–426. He actually proved (3) with $\mathscr{K}_n(P^n,c,\gamma_n)$ replaced by $\mathscr{H}_n(P^n,c)$ and without requiring (2). The assumption (2) is an assumption on the fixed sequence $\{P^n\}$ of probability measures. Beran's proof proceeds as follows:

Under (i) and (iii), there exists a sequence of probability measures $\{Q^n(\mathbf{h})\}$ such that for every $0 < b < \infty$,

(4) $$\sup_{\|\mathbf{h}\|\leq b} \Sigma_i \int \{q_{ni}^{1/2}(\mathbf{h}) - p_{ni}^{1/2} - (1/2)\,\mathbf{h}'\xi_{ni}\, p_{ni}^{1/2}\}^2 d\nu_{ni} = o(1).$$

Consequently,

(5) $$\lim_n \sup_{\|\mathbf{h}\|\leq b} \Sigma_i \int \{q_{ni}^{1/2}(\mathbf{h}) - p_{ni}^{1/2}\}^2 d\nu_{ni} = 4^{-1} b^2,$$

and for n sufficiently large, the family $\{Q^n(\mathbf{h}),\ \|\mathbf{h}\| \leq b,\ \mathbf{h}\in\mathbb{R}^p\}$ is a subset of $\mathscr{H}_n(P^n,(b/2))$. Hence, $\forall\ c > 0$, $\forall$ sequence of statistics $\{\hat{S}_n\}$,

(6) $$\liminf_n \inf_{\hat{S}_n} \sup_{Q^n\in \mathscr{H}_n(P^n,c)} R_n(\hat{S}_n,Q^n)$$
$$\geq \liminf_n \inf_{\hat{S}_n} \sup_{\|\mathbf{h}\|\leq 2c} R_n(\hat{S}_n,Q^n(\mathbf{h})).$$

Then the proof proceeds as in Hájek – Le Cam setup for the parametric family $\{Q^n(\mathbf{h}),\ \|\mathbf{h}\| \leq b\}$, under the l.a.n. property of the family $\{Q^n(\mathbf{h}),\ \|\mathbf{h}\| \leq b\}$ with $b = 2c$, which is implied by (4).

Thus (3) would be proved if we verify (6) with $\mathscr{H}_n(P^n,c)$ replaced by $\mathscr{K}_n(P^n,c,\eta_n)$, under the additional assumption (2). That is, we have to show that there exist sequences $0 < \eta_{n1} \to 0$, $0 < \eta_{n2} \to 0$ such that the above family $\{Q^n(\mathbf{h}),\ \|\mathbf{h}\| \leq b\}$ is a subset of $\mathscr{K}_n(P^n,(b/2),\eta_n)$ for sufficiently large n. To that effect we recall the family $\{Q^n(\mathbf{h})\}$ from Beran. With ξ_{ni} as in (i) – (iii), let ξ_{nij} denote the jth component of ξ_{ni}, $1 \leq j \leq p$, $1 \leq i \leq n$. By (iii) there exist a sequence $\epsilon_n > 0$, $\epsilon_n \downarrow 0$ such that

$$\max_{1\leq j\leq p} \sum_i \int \xi_{nij}^2\, I(|\xi_{nij}| > \epsilon_n)\, dP_{ni} = o(1).$$

Now, define

$$\xi_{nij}^* := \xi_{nij}\, I(|\xi_{nij}| \le \epsilon_n), \qquad \bar{\xi}_{nij} := \xi_{nij}^* - \int \xi_{nij}^*\, dP_{ni}, \qquad 1 \le j \le p,$$

$$\bar{\xi}_{ni} := (\bar{\xi}_{ni1}, \ldots, \bar{\xi}_{nip})', \qquad 1 \le i \le n.$$

Note that

$$(7) \qquad \|\bar{\xi}_{ni}\| \le 2p\epsilon_n, \qquad \int \bar{\xi}_{ni} dP_{ni} = 0, \qquad 1 \le i \le n.$$

For a $0 < b < \infty$, $\|\mathbf{h}\| \le b$, $1 \le i \le n$, define

$$\begin{aligned} q_{ni}(\mathbf{h}) &= (1 + \mathbf{h}'\bar{\xi}_{ni})p_{ni}, & \epsilon_n < (2bp)^{-1}, \\ &= p_{ni}, & \epsilon_n \ge (2bp)^{-1}. \end{aligned}$$

Because of (7), $\{q_{ni}(\mathbf{h}), \|\mathbf{h}\| \le b, 1 \le i \le n\}$ are probability density functions. Let $\{Q_{ni}(\mathbf{h}); \|\mathbf{h}\| \le b, 1 \le i \le n\}$ denote the corresponding probability measures and $Q^n(\mathbf{h}) = Q_{n1}(\mathbf{h})x....xQ_{nn}(\mathbf{h})$.

Now, note that for $\|\mathbf{h}\| \le b$, $1 \le i \le n$,

$$\begin{aligned} \int (q_{ni}(\mathbf{h}) - p_{ni})^2\, d\mu_{ni} &= 0, & \epsilon_n \ge (2bp)^{-1}, \\ &= \int (\mathbf{h}'\bar{\xi}_{ni})^2 p_{ni}^2\, d\mu_{ni}, & \epsilon_n < (2bp)^{-1}. \end{aligned}$$

Consequently, since $\epsilon_n \downarrow 0$, $\epsilon_n < (2bp)^{-1}$ eventually, and

$$\sup_{\|\mathbf{h}\| \le b} \max_i \int (q_{ni}(\mathbf{h}) - p_{ni})^2\, d\mu_{ni} \le (2p\epsilon_n)^2\, b^2 \max_i \int p_{ni}^2\, d\mu_{ni} =: \eta_{n1}.$$

Similarly, for a sufficiently large n,

$$\sup_{\|\mathbf{h}\| \le b} \max_i \int (q_{ni}^{1/2}(\mathbf{h}) - p_{ni}^{1/2})^2\, d\nu_{ni} \le 2bp\epsilon_n =: \eta_{n2}, \quad \text{say}.$$

Because of (2) and because $\epsilon_n \downarrow 0$, $\max\{\eta_{n1}, \eta_{n2}\} \longrightarrow 0$.

Consequently, for every $b > 0$ and for n sufficiently large, $\{Q^n(\mathbf{h}), \|\mathbf{h}\| \le b\}$ is a subset of $\mathscr{K}_n(P^n,(b/2),\eta_n)$ with the above η_{n1}, η_{n2} and an analogue of (6) with $\mathscr{H}_n(P^n,c)$ replaced by $\mathscr{K}_n(P^n,(b/2),\eta_n)$ holds. The rest is the same as in Beran. □

We shall now show that $\boldsymbol{\beta}^+$ achieves the lower bound in (3). Fix a $\boldsymbol{\beta} \in \mathbb{R}^p$ and consider the model (1.1.1). As before, let F_{ni} be the actual d.f. of e_{ni}, $1 \le i \le n$, and suppose we model the errors to be i.i.d F, F symmetric around zero. The d.f. F need not be known. Then the actual and the

modeled d.f. of Y_{ni} of (1.1.1) is $F_{ni}(\cdot - \mathbf{x}_{ni}'\beta)$, $F(\cdot - \mathbf{x}_{ni}'\beta)$, respectively.

In Theorem 5.6c.1 take $X_{ni} \equiv Y_{ni}$ and $\{Q_{ni}, P_{ni}, \nu_{ni}\}$ as follows:

$$(8) \qquad Q_{ni}^{\beta}(Y_{ni} \le \cdot) = F_{ni}(\cdot - \mathbf{x}_{ni}'\beta), \ P_{ni}^{\beta}(Y_{ni} \le \cdot) = F(\cdot - \mathbf{x}_{ni}'\beta),$$
$$\mu_{ni}^{\beta}(\cdot) = G(\cdot - \mathbf{x}_{ni}'\beta), \qquad \nu_{ni} \equiv \lambda, \qquad 1 \le i \le n.$$

Also, let $Q_{\beta}^{n} = Q_{n1}^{\beta} x \dots x Q_{nn}^{\beta}$; $P_{\beta}^{n} = P_{n1}^{\beta} x \dots x P_{nn}^{\beta}$. The absence of β from the sub– or the super– script of a probability measure indicates that the measure is being evaluated at $\beta = 0$. Thus, for example we write Q^n for Q_0^n $(= \prod_{i=1}^{n} F_{ni})$ and P^n for P_0^n, etc. Also for an integrable function g write $\int g$ for $\int g \, d\lambda$.

Let f_{ni}, f denote the respective densities of F_{ni}, F, w.r.t. λ. Then $q_{ni}^{\beta}(\cdot) = f_{ni}(\cdot - \mathbf{x}_{ni}'\beta)$, $p_{ni}^{\beta}(\cdot) = f(\cdot - \mathbf{x}_{ni}'\beta)$ and, because of the translation invariance of the Lebesgue measure,

$$(9) \qquad \mathscr{H}_n(P_{\beta}^n, c) = \{Q_{\beta}^n \in \Pi^n \, ; \, \Sigma_i \int \{(q_{ni}^{\beta})^{1/2} - (p_{ni}^{\beta})^{1/2}\}^2 \le c^2\}$$
$$= \{Q^n \in \Pi^n \, ; \, \Sigma_i \int (f_{ni}^{1/2} - f^{1/2})^2 \le c^2\} = \mathscr{H}_n(P^n, c).$$

That is the set $\mathscr{H}_n(P_{\beta}^n, c)$ does not depend on β. Similarly,

$$\mathscr{K}_n(P_{\beta}^n, c, \eta_n) = \{Q^n \in \Pi^n; \, Q^n \in \mathscr{H}_n(P^n, c), \, \max_i \int (f_{ni} - f)^2 \, dG \le \eta_{n1},$$
$$\max_i \int (f_{ni}^{1/2} - f^{1/2})^2 \le \eta_{n2}\} = \mathscr{K}_n(P^n, c, \eta_n).$$

Next we need to define the relevant functionals. For $\mathbf{t} \in \mathbb{R}^p$, $y \in \mathbb{R}$, $1 \le i \le n$, define

$$(10) \qquad m_{ni}^{+}(y, \mathbf{t}) = F_{ni}(y + \mathbf{x}_{ni}'(\mathbf{t} - \beta)) - 1 + F_{ni}(-y + \mathbf{x}_{ni}'(\mathbf{t} - \beta)),$$
$$\mathbf{b}_n(y, \mathbf{t}) := \Sigma_i \, \mathbf{A}\mathbf{x}_{ni} \, m_{ni}^{+}(y, \mathbf{t}),$$
$$\mu_n(\mathbf{t}, Q_{\beta}^n) \equiv \mu_n(\mathbf{t}, \mathbf{F}) := \int \|\mathbf{b}_n(y, \mathbf{t})\|^2 \, dG(y),$$
$$\mathbf{F}' := (F_{n1}, \cdots, F_{nn}).$$

Now, recall the definition of ψ from (5.6a.2) and let $\mathbf{T}_n(\beta, Q_{\beta}^n) \equiv \mathbf{T}_n(\beta, \mathbf{F})$ be defined by the relation

(11) $\mathbf{T}_n(\boldsymbol{\beta}, \mathbf{F}) := \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X}\int f^2 dG)^{-1}\int \Sigma_i\, \mathbf{x}_{ni}[F_{ni}(y) - 1 + F_{ni}(-y)]\, d\psi(y).$

Note that, with $\mathbf{b}_n(y) \equiv \mathbf{b}_n(y, \boldsymbol{\beta})$,

(12) $$\mathbf{A}^{-1}(\mathbf{T}_n(\boldsymbol{\beta}, \mathbf{F}) - \boldsymbol{\beta}) = (\int f^2 dG)^{-1} \int \mathbf{b}_n(y)\, d\psi(y).$$

Some times we shall write $\mathbf{T}_n(\mathbf{F})$ for $\mathbf{T}_n(\boldsymbol{\beta}, \mathbf{F})$.

Observe that if $\{F_{ni}\}$ are symmetric around 0, then $\mathbf{T}_n(\boldsymbol{\beta}, \mathbf{F}) = \boldsymbol{\beta} = \mathbf{T}_n(\boldsymbol{\beta}, P^n_{\boldsymbol{\beta}})$. In general, the quantity $\mathbf{A}^{-1}(\mathbf{T}_n(\mathbf{F}) - \boldsymbol{\beta})$ measures the asymptotic bias in $\boldsymbol{\beta}^+$ due to the asymmetry of the errors.

We shall prove the l.a.m. property of $\boldsymbol{\beta}^+$ by showing that $\mathbf{T}_n$ is H-differentiable and that $\boldsymbol{\beta}^+$ is an estimator of $\mathbf{T}_n$ that achieves the lower bound in (3). To that effect we first state a lemma. Its proof follows from Theorem 5.5.4 in the same fashion as that of Lemma 5.6b.1 and Corollary 5.6b.2 from Theorem 5.5.3. Observe that the conditions (5.5.35) and (5.5.11$^+$) with $\mathbf{D} = \mathbf{XA}$, respectively, become

(13) $$\int \|\mathbf{b}_n(y)\|^2\, dG(y) = O(1),$$

(14) $$\liminf_n \inf_{\|\boldsymbol{\theta}\|=1} \boldsymbol{\theta}'\mathbf{A}\mathbf{X}'\int \boldsymbol{\Lambda}^+ g\,dG\, \mathbf{X}\mathbf{A}\,\boldsymbol{\theta} \geq \alpha, \qquad \text{for an } \alpha > 0,$$

where $\boldsymbol{\Lambda}^+$ is defined at (5.5.38) and g is as in (5.6a.3).

Lemma 5.6c.1. *Assume that* (1.1.1) *holds with the actual d.f.'s of* $\{e_{ni}, 1 \leq i \leq n\}$ *equal to* $\{F_{ni}, 1 \leq i \leq n\}$ *and suppose that we model the errors to be i.i.d* F, F *symmetric around zero. In addition, assume that* (5.3.8); (5.5.1), (5.5.3), (5.5.4), (5.5.6), (5.5.7), (5.5.9) *with* $\mathbf{D} = \mathbf{XA}$, $G_n \equiv G$; (5.5.68), (5.6a.13), (5.6b.6) *and* (13) *hold. Then* (5.5.8) *and its variant where the argument* y *in the integrand is replaced by* $-y$, (5.5.33), (14) *and the following hold.*

(15) $$\mathbf{A}^{-1}(\boldsymbol{\beta}^+ - \mathbf{T}_n(\mathbf{F})) = -\{2\int f^2\, dG\}^{-1}\, \mathbf{Z}_n^+ + o_p(1), \quad \textit{under } \{Q^n\}.$$

where

(16) $$\mathbf{Z}_n^+ = \Sigma_i\, \mathbf{A}\mathbf{x}_{ni}\, \{\psi(-e_{ni}) - \psi(e_{ni}) - \int \mathbf{m}_{ni}^+(y)\, dG(y)\},$$

with $\mathbf{m}_{ni}^+(y) \equiv \mathbf{m}_{ni}^+(y, \boldsymbol{\beta})$ *and* ψ *as in* (5.6a.2). □

Now, define, for an $0 < a < \infty$,

$$\mathscr{M}_n(P^n, a) = \{Q^n \in \Pi^n;\ Q^n = \prod_{i=1}^{n} F_{ni},\ \max_i \int |f_{ni} - f|^r\, dG \longrightarrow 0,\ r = 1, 2,$$
$$\max_i \|F_{ni} - F\|_\infty \longrightarrow 0,\ \int [\Sigma_i \|\mathbf{A}\mathbf{x}_{ni}\|\ |F_{ni} - F|]^2\, dG \le a^2\}.$$

Lemma 5.6c.2. *Assume that* (1.1.1) *holds with the actual d.f.'s of* $\{e_{ni}, 1 \le i \le n\}$ *equal to* $\{F_{ni}, 1 \le i \le n\}$ *and suppose that we model the errors to be i.i.d* F, F *symmetric around zero. In addition, assume that* (5.3.8), (5.5.1), (5.5.68) *and the following hold.*

(17) G *is a finite measure.*

Then, for every $0 < a < \infty$ *and sufficiently large* n,

$$\mathscr{M}_n(P^n, a) \supset \mathscr{K}_n(P^n, b_a, \eta_n), \qquad b_a := (4p\alpha)^{-1/2} a, \quad \alpha := G(\mathbb{R}).$$

Moreover, all assumptions of Lemma 5.6c.1 *are satisfied.*

Proof. Fix an $0 < a < \infty$. It suffices to show that

(19) $$\Sigma_i \int (f_{ni}^{1/2} - f^{1/2})^2 \le b_a^2, \qquad n \ge 1,$$

and

(20) $$\text{(a)}\ \max_i \int (f_{ni} - f)^2\, dG \le \eta_{n1}, \qquad n \ge 1,$$

$$\text{(b)}\ \max_i \int (f_{ni}^{1/2} - f^{1/2})^2 \le \eta_{n2}, \qquad n \ge 1,$$

imply all the conditions describing $\mathscr{M}_n(P^n, a)$.

Claim: (19) implies $\int [\Sigma_i \|\mathbf{A}\mathbf{x}_{ni}\|\ |F_{ni} - F|]^2\, dG \le a^2, \qquad n \ge 1.$

By the C–S inequality,

(21) $$|F_{ni}(x) - F(x)|^2 = |\int_{-\infty}^{x} (f_{ni} - f)|^2$$
$$\le \int_{-\infty}^{x} (f_{ni}^{1/2} - f^{1/2})^2 \cdot \int_{-\infty}^{x} (f_{ni}^{1/2} + f^{1/2})^2$$
$$\le 4 \int (f_{ni}^{1/2} - f^{1/2})^2, \qquad 1 \le i \le n,\ x \in \mathbb{R}.$$

Hence,

$$\int [\Sigma_i \|\mathbf{A}\mathbf{x}_{ni}\|\ |F_{ni} - F|]^2\, dG \le \Sigma_i \|\mathbf{A}\mathbf{x}_{ni}\|^2 \cdot \Sigma_i \int (F_{ni} - F)^2\, dG$$
$$\le 4pa^2 \cdot \Sigma_i \int (f_{ni}^{1/2} - f^{1/2})^2,$$

which proves the Claim.

The finiteness of G together with (21) and (20b) with $\eta_{n2} \longrightarrow 0$ imply that $\max_i \|F_{ni} - F\|_\infty \longrightarrow 0$ in a routine fashion. The rest uses (5.5.66), (5.5.67) and details are straightforward. □

Now let $\varphi(y) = \psi(-y) - \psi(y)$, $y \in \mathbb{R}$. Note that $d\psi(-y) \equiv -d\psi(y)$, $d\varphi \equiv -2\, d\psi$, $d\psi = f dG$ and because F is symmetric around 0, $\int \varphi f = 0$. Let

$$\sigma^2 = \operatorname{Var}\{\psi(e) \mid F\}, \quad \tau = \int f^2 dG, \quad \rho = (\varphi/\sigma),$$

$$\xi_{ni} \equiv \xi_{ni}(Y_{ni}, \beta) \equiv \mathbf{A}\mathbf{x}_{ni}\, \rho(e_{ni}).$$

Use the above facts to obtain

$$\begin{aligned}
2\, \Sigma_i \int \xi_{ni}(y, \beta)\, (p_{ni}^{\beta}(y))^{1/2} \{(q_{ni}^{\beta}(y))^{1/2} - (p_{ni}^{\beta}(y))^{1/2}\}^2\, dy \\
&= 2\, \Sigma_i\, \mathbf{A}\mathbf{x}_{ni} \int \rho\, f^{1/2}\, (f_{ni}^{1/2} - f^{1/2}) \\
&= \Sigma_i\, \mathbf{A}\mathbf{x}_{ni} \{ \int \rho\, f_{ni} - \int \rho\, (f_{ni}^{1/2} - f^{1/2})^2 \} \\
&= -\sigma^{-1}\, \Sigma_i\, \mathbf{A}\mathbf{x}_{ni} \{ \int [F_{ni} - F]\, d\varphi - \int \rho\, (f_{ni}^{1/2} - f^{1/2})^2 \}\, . \\
(22) \qquad &= \sigma^{-1} \Sigma_i\, \mathbf{A}\mathbf{x}_{ni} \{ 2 \int [F_{ni} - F]\, f dG - \int \rho\, (f_{ni}^{1/2} - f^{1/2})^2 \}.
\end{aligned}$$

The last but one equality follows from integrating the first term by parts.

Now consider the r.h.s. of (12). Note that because F and G are symmetric around 0,

$$\begin{aligned}
\int \mathbf{b}_n\, f dG &= \int \Sigma_i\, \mathbf{A}\mathbf{x}_{ni}\, [F_{ni}(y) - 1 + F_{ni}(-y)]\, d\psi(y) \\
&= \int \Sigma_i\, \mathbf{A}\mathbf{x}_{ni}\, [F_{ni}(y) - F(y) + F_{ni}(-y) - F(-y)]\, d\psi(y) \\
(23) \qquad &= 2 \int \Sigma_i\, \mathbf{A}\mathbf{x}_{ni}\, [F_{ni} - F]\, f dG.
\end{aligned}$$

Recall that by definition $\mathbf{T}_n(\beta, P_\beta^n) \equiv \beta$. Now take $\mathcal{A}_n$ of (ii) of the H–differentiable requirement to be $\mathbf{A}^{-1} \tau \sigma^{-1}$ and conclude from (18), (22), (23), that

$$\begin{aligned}
\| \mathcal{A}_n \{ \mathbf{T}_n(\beta, Q_\beta^n) - \mathbf{T}_n(\beta, P_\beta^n) \} - \\
- 2 \Sigma_i \int \xi_{ni}(y, \beta) (p_{ni}^{\beta}(y))^{1/2} \{(q_{ni}^{\beta}(y))^{1/2} - (p_{ni}^{\beta}(y))^{1/2}\}^2\, dy\, \|
\end{aligned}$$

$$\leq \|\Sigma_i \mathbf{A}\mathbf{x}_{ni} \int \int \rho \, (f_{ni}^{1/2} - f^{1/2})^2\| \leq \max_i \|\mathbf{A}\mathbf{x}_{ni}\| \cdot \|\rho\|_\infty \cdot b_a^2 = o(1),$$

uniformly for $\{Q^n\} \in \mathscr{K}_n(P^n, b_a, \gamma_n)$.

This proves that the requirement (ii) of the Definition 5.6c.1 is satisfied by the functional $\mathbf{T}_n$ with the $\{\xi_{ni}\}$ given as above. The fact that these $\{\xi_{ni}\}$ satisfy (i) and (iii) of the Definition 5.6c.1 follows from (5.3.8), (5.5.1), (17), (18) and the symmetry of F. This then verifies the H-differentiability of the above m.d. functional $\mathbf{T}_n$.

We shall now derive the asymptotic distribution of $\boldsymbol{\beta}^+$ under any sequence $\{Q^n\} \in \mathscr{M}_n(P^n, a)$, under the conditions of Lemma 5.6c.2. For that reason consider $\mathbf{Z}_n^+$ of (16). Note that under Q^n, $(1/2)\mathbf{Z}_n^+$ is the sum of independent centered triangular random arrays and the boundedness of ψ and (5.5.1), imply, via the L–F CLT, that $\mathbf{C}_n^{-1/2}\, \mathbf{Z}_n^+ \xrightarrow[d]{} N(0, \mathbf{I}_{p\times p})$, where

$$\mathbf{C}_n = 4^{-1} E\, \mathbf{Z}_n^+ \mathbf{Z}_n^{+\prime} = \Sigma_i \mathbf{A}\mathbf{x}_{ni}\mathbf{x}_{ni}'\mathbf{A}\, \sigma_{ni}^2, \quad \sigma_{ni}^2 = \mathrm{Var}\{\psi(e_{ni}) \,|\, F_{ni}\},\ 1 \leq i \leq n.$$

But the boundedness of ψ implies that $\max_i |\sigma_{ni}^2 - \sigma^2| \to 0$, for every $Q^n \in \mathscr{M}_n(P^n, a)$, where $\sigma^2 = \mathrm{Var}\,\{\psi(e_1) | F\}$. Therefore $\sigma^{-1}\mathbf{Z}_n^+ \xrightarrow[d]{} N(0, \mathbf{I}_{p\times p})$.

Consequently, from (15),

$$\lim_{c\to 0} \lim_n \sup_{\boldsymbol{\beta}} \sup_{Q^n \in \mathscr{K}_n(P^n_{\boldsymbol{\beta}}, c, \eta_n)} E\{\, \mathscr{U}(\|A_n(\boldsymbol{\beta}^+ - \mathbf{T}_n(\boldsymbol{\beta}, Q^n_{\boldsymbol{\beta}}))\| \,|\, Q^n_{\boldsymbol{\beta}}\} = E\, \mathscr{U}(\|\mathbf{Z}\|).$$

for every bounded nondecreasing function $\mathscr{U}$, where $\mathbf{Z}$ is a $N(0, \mathbf{I}_{p\times p})$ r. v..

This and Lemma 5.6c.2 shows that the sequence of the m.d. estimators $\{\boldsymbol{\beta}^+\}$ achieves the lower bound of (3) and hence is l.a.m. □

Remark 5.6c.1. It is an *interesting problem* to see if one can remove the requirement of the finiteness of the integrating measure G in the above l.a.m. result. The l.a.m. property of $\{\hat{\boldsymbol{\beta}}\}$ can be obtained in a similar fashion. For an alternative definition of l.a.m. see Millar (1984) where, among other things, he proves the l.a.m. property, in his sense, of $\{\hat{\boldsymbol{\beta}}\}$ for $p = 1$.

A problem: To this date an appropriate extension of Beran (1978) to the model (1.1.1) does not seem to be available. Such an extension would provide asymptotically fully efficient estimators at every symmetric density with finite Fisher information and would also be l.a.m. □

Note: The contents of this chapter are based on the works of Williamson (1979, 1982), Koul (1979, 1980, 1984, 1985a,b), Koul and DeWet (1983), Basaw and Koul (1988) and Dhar (1991a, b). □□

CHAPTER 6

GOODNESS–OF–FIT TESTS FOR THE ERRORS

6.1. INTRODUCTION

Consider the model (1.1.1) and the goodness–of–fit hypothesis

(1) $$H_0\colon \mathrm{F}_{ni} \equiv \mathrm{F}_0, \qquad \mathrm{F}_0 \text{ a known continuous d.f..}$$

This is a classical problem yet not much is readily available in literature. Observe that even if F_0 is known, having an unknown $\boldsymbol{\beta}$ in the model poses a problem in constructing tests of H_0 that would be implementable, at least asymptotically.

One test of H_0 could be based on $\hat{\mathrm{D}}_1$ of (1.3.3). This test statistic is suggested by looking at the estimated residuals and mimicking the one sample location model technique. In general, its large sample distribution depends on the design matrix. In addition, it does not reduce to the Kiefer (1959) tests of goodness–of–fit in the k–sample location problem when (1.1.1) is reduced to this model. The test statistics that overcome these deficiencies are those that are based on the w.e.p.'s $\mathbf{V}$ of (1.1.2). For example, the two candidates that will be considered in this chapter are

(2) $$\hat{\mathrm{D}}_2 := \sup_y |\mathbf{W}^0(y, \hat{\boldsymbol{\beta}})|, \qquad \hat{\mathrm{D}}_3 := \sup_y \|\mathbf{W}^0(y, \hat{\boldsymbol{\beta}})\|,$$

where $\hat{\boldsymbol{\beta}}$ is an estimator of $\boldsymbol{\beta}$ and,

(3) $$\mathbf{W}^0(y, \mathbf{t}) := (\mathbf{X}'\mathbf{X})^{-1/2}\{\mathbf{V}(y, \mathbf{t}) - \mathbf{X}'\mathbf{1}\,\mathrm{F}_0(y)\}, \qquad y \in \mathbb{R},\ \mathbf{t} \in \mathbb{R}^p,$$

$$\mathbf{1}' := (1, \dots, 1)_{1\times n}.$$

Other classes of tests are based on $\mathrm{K}^0_{\mathbf{X}}(\hat{\boldsymbol{\beta}}_{\mathbf{X}})$ and $\inf\{\mathrm{K}^0_{\mathbf{X}}(\mathbf{t}), \mathbf{t}\in\mathbb{R}^p\}$, where $\mathrm{K}^0_{\mathbf{X}}$ is equals to the $\mathrm{K}_{\mathbf{X}}$ of (1.3.2) with $\mathbf{W}$ replaced by $\mathbf{W}^0$ in there.

Section 6.2a discusses the asymptotic null distributions (a.n.d.'s) of the supremum distance test statistics for H_0 when $\boldsymbol{\beta}$ is estimated arbitrarily and asymptotically efficiently. Also discussed in this section are some asymptotically distribution free (a.d.f.) tests for H_0. Some comments about the asymptotic power of these tests appear at the end of this section. Section 6.2b discusses a smooth bootstrap distribution of $\hat{\mathrm{D}}_3$.

Analogous results for tests of H_0 based on L_2–distances involving the ordinary and weighted empirical processes appear in Section 6.3.

A closely related problem to H_0 is that of testing the composite hypothesis

(4) $\qquad H_1: \ F_{ni}(\cdot) = F_0(\cdot/\sigma), \ \sigma > 0, \ F_0$ a known d.f..

Modifications of various tests of H_0 and their asymptotic null distributions are discussed in Section 6.4.

Another problem of interest is to test the composite hypothesis of symmetry of the errors:

(5) $\qquad H_s$: $F_{ni} = F, \ 1 \le i \le n, n \ge 1$; F a d.f. symmetric around 0.

This is a more general hypothesis than H_0. In some situations it may be of interest to test H_s before testing, say, that the errors are normally distributed. Rejection of H_s would *a priori* exclude any possibility of normality of the errors. A test of H_s could be based on

(6) $$\hat{D}_{1s} := \sup_y |W_1^+(y, \hat{\beta})|,$$

where

(7) $$W_1^+(y, \mathbf{t}) := n^{-1/2} \sum_{i=1}^{n} [I(Y_{ni} \le y + \mathbf{x}_{ni}'\mathbf{t}) - I(-Y_{ni} < y - \mathbf{x}_{ni}'\mathbf{t})]$$
$$:= H_n(y, \mathbf{t}) - 1 + H_n(-y, \mathbf{t}), \qquad y \in \mathbb{R}, y \in \mathbb{R}^p,$$

with H_n as in (1.2.1). Other candidates are

(8) $$\hat{D}_{2s} := \sup_y |\mathbf{W}^+(y, \hat{\beta})|,$$
$$\hat{D}_{3s} := \sup_y \|\mathbf{W}^+(y, \hat{\beta})\| = \sup_y[\mathbf{V}^{+\prime}(y, \hat{\beta})(\mathbf{X}'\mathbf{X})^{-1}\mathbf{V}^+(y, \hat{\beta})]^{1/2},$$

where

(9) $$\mathbf{W}^+ := \mathbf{A}\mathbf{V}^+, \qquad \mathbf{V}^{+\prime} := (V_1^+, \dots, V_p^+), \text{ with}$$
$$V_j^+(y, \mathbf{t}) := V_j(y, \mathbf{t}) - \sum_{i=1}^{n} x_{nij} + V_j(-y, \mathbf{t}), \qquad 1 \le j \le p, \ \ y \in \mathbb{R}, \ \ \mathbf{t} \in \mathbb{R}^p.$$

Yet other tests can be obtained by considering various L_2-norms involving W_1^+ and $\mathbf{W}^+$. The asymptotic null distribution of all of these test statistics is given in Section 6.5.

It will be observed that the tests based on the vectors $\mathbf{W}^0$ and $\mathbf{W}^+$ of w.e.p.'s will have asymptotic distributions similar to their counterparts in the k–sample location models. Consequently these tests can use, at least for the large samples, the null distribution tables that are available for such problems. For the sake of the completeness some of these table are reproduced in the following sections.

6.2. THE SUPREMUM DISTANCE TESTS

6.2a. Asymptotic Null Distributions.

To begin with, define, for $0 \le t \le 1$, $s \in \mathbb{R}^p$,

$$(1) \qquad W_1(t, s) := n^{1/2}\{H_n(F_0^{-1}(t), s) - t\}, \qquad \mathbf{W}(t, \mathbf{s}) := \mathbf{W}^o(F_0^{-1}(t), \mathbf{s}).$$

Let

$$(2) \qquad \hat{W}_1(t) := W_1(t, \hat{\beta}), \qquad \hat{\mathbf{W}}(t) := \mathbf{W}(t, \hat{\beta}), \qquad 0 \le t \le 1.$$

Clearly, if F_0 is continuous then the distribution of $\hat{D}_j$, $j = 1, 2, 3$, is the same as that of $\|\hat{W}_1\|_\infty$, $\sup\{|\hat{\mathbf{W}}(t)|;\ 0 \le t \le 1\}$, $\sup\{\|\hat{\mathbf{W}}(t)\|;\ 0 \le t \le 1\}$, respectively. Consequently, from Corollaries 2.3.3 and 2.3.5 one readily obtains the following Theorem 6.2a.1. Recall the conditions ($\mathbf{F}_0\mathbf{1}$) and (**NX**) from Corollary 2.3.1 and just after Corollary 2.3.2.

Theorem 6.2a.1. *Suppose that the model* (1.1.1) *and* H_0 *hold. In addition, assume that* X *and* F_0 *satisfy* (**NX**) *and* ($\mathbf{F}_0\mathbf{1}$), *and that* $\hat{\beta}$ *satisfies*

$$(3) \qquad \|A^{-1}(\hat{\beta} - \beta)\| = O_p(1).$$

Then

$$(4) \qquad \sup|W_1(t, \hat{\beta}) - \{W_1(t, \beta) + q_0(t)\cdot n^{1/2}\, \bar{x}_n' A \cdot A^{-1}(\hat{\beta} - \beta)\}| = o_p(1),$$

$$(5) \qquad \sup\|\mathbf{W}(t, \hat{\beta}) - \{\mathbf{W}(t, \beta) + q_0(t)\cdot A^{-1}(\hat{\beta} - \beta)\}\| = o_p(1),$$

where $q_0 := f_0(F_0^{-1})$ *and the supremum is over* $0 \le t \le 1$. □

Write $W_1(t)$, $\mathbf{W}(t)$ for $W_1(t, \beta)$, $\mathbf{W}(t, \beta)$, respectively. The following corollary gives the weak limits of W_1 and $\mathbf{W}$ under H_0.

Lemma 6.2a.2. *Suppose that the model* (1.1.1) *and* H_0 *hold. Then*

$$(7) \qquad W_1 \Rightarrow B, \ B \text{ a Brownian bridge in } \mathbb{C}[0, 1].$$

In addition, if X *satisfies* (**NX**), *then,*

$$(8) \qquad \mathbf{W} \Rightarrow B' := (B_1, \ldots., B_p)$$

where $B_1, \ldots., B_p$ *are independent Brownian bridges in* $\mathbb{C}[0, 1]$.

Proof. The result (7) is well known or may be deduced from Corollary 2.2a.2. The same corollary implies (8). To see this, rewrite

(9) $$\mathbf{W}(t) = \mathbf{A}\sum_i \mathbf{x}_{ni}\{I(e_{ni} \le F_0^{-1}(t)) - t\} = \mathbf{A}\mathbf{X}'\alpha_n(t),$$

where $\alpha_n(t) := (\alpha_{n1}(t), \ldots, \alpha_{nn}(t))'$, with

$$\alpha_{ni}(t) := \{I(e_{ni} \le F_0^{-1}(t)) - t\}, \qquad 1 \le i \le n,\ 0 \le t \le 1.$$

Clearly, under H_0,

(10) $$E\mathbf{W} \equiv 0, \qquad \mathrm{Cov}(\mathbf{W}(s), \mathbf{W}(t)) = (s \wedge t - st)\mathbf{I}_{p \times p}, \qquad 0 \le s, t \le 1.$$

Now apply Corollary 2.2a.2 p times, jth time to the w.e.p. with the weights and r.v.'s given as in (11) below, $1 \le j \le p$, to conclude (8).

(11) weights $\mathbf{d}_{(j)} \equiv$ the j^{th} column of $\mathbf{XA}$, the r.v.'s $X_{ni} \equiv e_{ni}$, and $F \equiv F_0$, $1 \le j \le p$,

See (2.3.33) and (2.3.34) for ensuring the applicability of Corollary 2.2a.2 to this case. □

Remark 6.2a.1. From (5) it follows that if $\hat{\beta}$ is chosen so that the finite dimensional asymptotic distributions of $\{\mathbf{W}(t) + q_0(t)\ \mathbf{A}^{-1}(\hat{\beta} - \beta);\ 0 \le t \le 1\}$ do not depend on the design matrix then the a.n.d.'s of $\hat{D}_j$, $j = 2, 3$, will also not depend on the design matrix. The classes of estimators that satisfy this requirment include M-, R- and m.d. estimators. Consequently, in these cases, the a.n.d.'s of $\hat{D}_j$, $j = 2, 3$, are design free.

On the other hand, from (4), the a.n.d. of $\hat{D}_1$ depends on the design matrix through $n^{1/2}\bar{\mathbf{x}}_n'\mathbf{A}$. Of course, if $\bar{\mathbf{x}}_n$ equals to zero, then this distribution is free from F_0 and the design matrix. □

Remark 6.2a.2. *The effect of estimating the parameter* β *efficiently.* To describe this, assume that

(12) F_0 has an a.c. density f_0 with a.e. derivative $\dot{f}_0$ satisfying
$$0 < I_0 := \int (\dot{f}_0/f_0)^2\, dF_0 < \infty.$$

Define

(13) $$s_{ni} := -\dot{f}_0(e_{ni})/f_0(e_{ni}), \quad 1 \le i \le n; \qquad \mathbf{s}_n := (s_{n1}, \ldots, s_{nn})',$$

and assume that the estimator $\hat{\beta}$ satisfies

(14) $$\mathbf{A}^{-1}(\hat{\beta} - \beta) = I_0^{-1}\,\mathbf{A}\mathbf{X}'\mathbf{s}_n + o_p(1).$$

Then, the approximating processes in (4) and (5), respectively, become

$$(15)\qquad \mathcal{W}_1(t) := W_1(t) + q_0(t)\cdot n^{1/2}\,\bar{x}_n' A\cdot \mathcal{I}_0^{-1}\, AX' s_n,$$

$$\mathcal{W}(t) := W(t) + q_0(t)\cdot \mathcal{I}_0^{-1}\, AX' s_n, \qquad 0 \le t \le 1.$$

Using the independence of the errors, one directly obtains

$$(16)\qquad E\,\mathcal{W}_1(s)\,\mathcal{W}_1(t) = \{s(1-t) - n\bar{x}_n'(X'X)^{-1}\bar{x}_n\, q_0(s)q_0(t)\mathcal{I}_0^{-1}\},$$

$$E\,\mathcal{W}(s)\,\mathcal{W}'(t) = \{s(1-t) - q_0(s)q_0(t)\mathcal{I}_0^{-1}\}\, I_{p\times p}, \qquad 0 \le s \le t \le 1.$$

The calculations in (16) use the facts that $Es_n \equiv 0$, $E\alpha_n(t)s_n' \equiv q_0(t)I_{n\times n}$.

From (16), Theorem 2.2a.1(i) applied to the quantities given in (11), and the uniform continuity of q_0, which is implied by (12), it readily follows that $\mathcal{W} \Rightarrow \mathbf{Z} := (Z_1, ..., Z_p)'$, where $Z_1, ..., Z_p$ are continuous independent Gaussian processes, each having the covariance function

$$(17)\qquad \rho(s, t) := s(1-t) - q_0(s)q_0(t)\mathcal{I}_0^{-1}, \qquad 0 \le s \le t \le 1.$$

Consequently,

$$(18)\qquad \hat{D}_2 \Rightarrow \sup\{|Z(t)|;\ 0\le t\le 1\}, \quad \hat{D}_3 \Rightarrow \sup\{\|Z(t)\|;\ 0\le t\le 1\}.$$

This shows that the a.n.d.'s of $\hat{D}_j$, $j = 2, 3$, are design free when an asymptotically efficient estimator of β is used in constructing the residuals while the same can not be said about $\hat{D}_1$.

Moreover, recall, say from Durbin (1975), that when testing for H_0 in the one sample location model, the Gaussian process Z_1 with the covariance function ρ appears as the limiting process for the analogue of $\hat{D}_1$. Note also that in this case, $\hat{D}_1 = \hat{D}_2 = \hat{D}_3$. However, it is the test based on $\hat{D}_3$ that provides the right extension of the one sample Kolmogorov goodness–of–fit test to the linear regression model (1.1.1) for testing H_0 in the sense that it includes the k–sample goodness–of–fit Kolmogorov type test of Kiefer (1959). That is, if we specialize (1.1.1) to the k–sample location model, then $\hat{D}_3$ reduces to the T_N' of Section 2 of Kiefer modulo the fact that we have to estimate β.

The distribution of $\sup\{|Z_1(t)|;\ 0\le t\le 1\}$ has been studied by Durbin (1976) when F_0 equals N(0, 1) and some other distributions. Consequently, one can use these results together with the independence of $Z_1, ..., Z_p$ to implement the tests based on $\hat{D}_2$, $\hat{D}_3$ in a routine fashion. □

Remark 6.2a.3. *Asymptotically distribution free (a.d.f.) tests.* Here we shall construct estimators of β such that the above tests become a.d.f. for testing H_0. To that effect, write $\mathbf{X}_n$ and $\mathbf{A}_n$ for $\mathbf{X}$ and $\mathbf{A}$ to emphasize their dependence on n. Recall that n is the number of rows in $\mathbf{X}_n$. Let m $= m_n$ be a sequence of positive integers, $m_n \le n$. Let $\mathbf{X}_m$ be $m_n \times p$ matrix obtained from some m_n rows of $\mathbf{X}_n$. A way to choose m_n and these rows will be discussed later on. Relable the rows of $\mathbf{X}_n$ so that its first m_n rows are the rows of $\mathbf{X}_m$ and let $\{e_{ni}^*,\ 1\le i\le m_n\}$, $\{Y_{ni}^*;\ 1\le i\le m_n\}$ denote the corresponding errors and observations, respectively. Define

(19) $$s_{ni}^* := -\dot{f}_0(e_{ni}^*)/f_0(e_{ni}^*),\ 1 \le i \le m_n; \quad \mathbf{s}_m^* := (s_{ni}^*,\ 1\le i\le m_n)',$$
$$\mathbf{T}_m := I_0^{-1}\, \mathbf{A}_m \mathbf{X}_m'\, \mathbf{s}_m^*, \qquad \mathbf{A}_m = (\mathbf{X}_m'\mathbf{X}_m)^{-1/2}.$$

Observe that under (12),

(20) $$E\mathbf{T}_m \equiv 0, \qquad E\mathbf{T}_m\mathbf{T}_m' \equiv I_0^{-1}\, \mathbf{I}_{p\times p}.$$

Consider the assumption

(21) $$m_n \le n,\ m_n \longrightarrow \infty \text{ such that}$$
$$(\mathbf{X}_n'\mathbf{X}_n)^{1/2}(\mathbf{X}_m'\mathbf{X}_m)^{-1}(\mathbf{X}_n'\mathbf{X}_n)^{1/2} \longrightarrow 2\mathbf{I}_{p\times p}.$$

The assumptions (21) and (NX) together imply

(22) $$\max_{1\le i\le m} \mathbf{x}_{ni}'\mathbf{A}_m\mathbf{A}_m\, \mathbf{x}_{ni} = o(1).$$

Consequently one obtains, with the aid of the Cramer–Wold LF–CLT, that

(23) $$\mathbf{T}_m \xrightarrow[d]{} N(0, I_0^{-1}\mathbf{I}_{p\times p}).$$

Now use $\{(\mathbf{x}_{ni}', Y_{ni}^*);\ 1 \le i \le m_n\}$ to construct an estimator $\hat{\beta}_m$ of β such that

(24) $$\mathbf{A}_m^{-1}(\hat{\beta}_m - \beta) = \mathbf{T}_m + o_p(1).$$

Note that, by (21) and (23), $\|\mathbf{A}_n^{-1}\mathbf{A}_m\|_\infty = O(1)$ and, hence

(25) $$\mathbf{A}_n^{-1}(\hat{\beta}_m - \beta) = \mathbf{A}_n^{-1}\mathbf{A}_m\mathbf{T}_m + o_p(1).$$

Therefore it follows that $\hat{\beta}_m$ satisfies (3). Define

$$\mathbf{K}^*(t) := \mathbf{W}(t) + \mathbf{A}_n^{-1}\mathbf{A}_m\mathbf{T}_m\, q_0(t), \qquad 0 \le t \le 1.$$

From (5) and (25) it now readily follows that

(26) $$\sup_{0\le t\le 1} \|\mathbf{W}(t, \hat{\beta}) - \mathbf{K}^*(t)\| = o_p(1).$$

We shall now show that

(27) $$\mathbf{K}^* \Rightarrow B, \text{ with } B \text{ as in (8).}$$

First, consider the covariance function of $\mathbf{K}^*$. By the independence of the errors and by (12) one obtains that

$$\begin{aligned} E\{I(e_{ni} \le F_0^{-1}(t)) - t\}\dot{f}_0(e_{nj}^*)/f_0(e_{nj}^*) &= 0, \quad i\ne j,\ 1\le i\le n,\ 1\le j\le m_n, \\ &= q_0(t), \quad 1\le i=j\le m_n, \quad 0\le t\le 1. \end{aligned}$$

Use this and direct calculations to obtain that

(28) $$\begin{aligned} E\mathbf{K}^*(s)\mathbf{K}^*(t) &= s(1-t)\mathbf{I}_{p\times p} \\ &\quad - I_0^{-1}q_0(s)q_0(t)[2\mathbf{I}_{p\times p} - (\mathbf{X}_n'\mathbf{X}_n)^{1/2}(\mathbf{X}_m'\mathbf{X}_m)^{-1}(\mathbf{X}_n'\mathbf{X}_n)^{1/2}], \\ &\qquad 0\le s\le t\le 1. \end{aligned}$$

Thus (21) implies that

(29) $$E\mathbf{K}^*(s)\mathbf{K}^*(t) \longrightarrow s(1-t)\mathbf{I}_{p\times p}, \qquad \forall\ 0\le s\le t\le 1.$$

Because of (8) and the uniform continuity of q_0, the relative compactness of the sequence $\{\mathbf{K}^*\}$ is *a priori* established, thereby completing the proof of (27). Consequently, we obtain the following

Corollary 6.2a.1. *Under* (1.1.1), H_0, (NX), (12), (21) *and* (24),

$$\hat{D}_{2m} \xrightarrow[d]{} \sup_{0\le t\le 1}\max_{1\le j\le p} |B_j(t)|, \qquad \hat{D}_{3m} \xrightarrow[d]{} \sup_{0\le t\le 1}\{\sum_{j=1}^{p} B_j^2(t)\}^{1/2},$$

where $\hat{D}_{jm}$ *stand for the* $\hat{D}_j$ *with* $\hat{\beta} = \hat{\beta}_m$, $j = 2, 3$. □

It thus follows, from the independence of the Brownian bridges $\{B_j,\ 1 \le j \le p\}$ and Theorem V.3.6.1 of Hajek and Sidak (1967), that the test that rejects H_0 when $\hat{D}_{2m} \ge d$ is of the asymptotic size α, provided d is determined from the relation

(30) $$2\sum_{j=1}^{\infty} (-1)^{j+1}e^{-2j^2d^2} = 1 - (1-\alpha)^{1/p}.$$

Let T_p stand for the the limiting r.v. of $\hat{D}_{3m}$. The distribution of T_p has been tabulated by Kiefer (1959) for $1 \le p \le 5$. Delong (1983) has also computed these tables for $1 \le p \le 7$. The following table is obtained from

Kiefer for $1 \leq p \leq 5$ and Delong for $p = 6, 7$, for the sake of completeness. The last place digit is rounded from their entries.

$\alpha \backslash p$	1	2	3	4	5	6	7
.001	1.9495	2.1516	2.3030	2.4301	2.5422	2.6437	2.7373
.005	1.7308	1.9417	2.0977	2.2280	2.3424	2.445	2.540
.01	1.6276	1.8427	2.0009	2.1326	2.2480	2.3525	2.4525
.02	1.5174	1.7370	1.8974	2.0305	2.1470	2.252	2.350
.025	1.480	1.702	1.8625	1.9961	2.116	2.217	2.315
.05	1.3581	1.5838	1.7473	1.8823	2.0001	2.1053	2.2031
.10	1.2239	1.4540	1.6196	1.7559	1.8746	1.981	2.0788
.15	1.1380	1.3703	1.5370	1.6740	1.7930	1.900	1.9977
.20	1.0728	1.3061	1.4734	1.6107	1.730	1.8352	1.9349
.25	1.0192	1.2530	1.4205	1.5579	1.6773	1.785	1.8825

Table 1: Values d *such that* $P(T_p \geq d) \simeq \alpha$ *for* $1 \leq p \leq 7$. *Obtained from Kiefer* (1959) *& Delong (personal communication).*

Note that for $p = 1$, $\hat{D}_{2m}$ and $\hat{D}_{3m}$ are the same tests and d of (30) is the same as the d of column 1 of Table 1 for various values of α.

The entries in Table 1 can be used to get the asymptotic critical level of $\hat{D}_{3m}$ for $1 \leq p \leq 7$. Thus for $p = 5$, $\alpha = .05$, the test that rejects H_0 when $\hat{D}_{3m} \geq 2.0001$ is of the asymptotic size .05, no matter what F_0 is within the class of d.f.'s satisfying (12).

Next, to make $\hat{D}_1$-test a.d.f., let $r = r_n$ be a sequence of positive integers, $r_n \leq n$, $r_n \to \infty$. Let X_r denote the $r_n \times p$ matrix obtain from some r_n rows of X_n. Relable the rows of X_n so that the first r_n rows are in X_r and let Y_i^o, e_i^o denote the corresponding Y_i's and e_i's. Let $A_r = (X_r' X_r)^{-1/2}$. Assume that

(31) (i) $\|n^{1/2}\bar{\mathbf{x}}_n'\mathbf{A}_r\| = O(1)$, and

(ii) $|n\bar{\mathbf{x}}_n(\mathbf{X}_r'\mathbf{X}_r)^{-1}\bar{\mathbf{x}}_n - 2r_n\bar{\mathbf{x}}_n'(\mathbf{X}_r'\mathbf{X}_r)^{-1}\bar{\mathbf{x}}_r| = o(1)$.

Let $\hat{\boldsymbol{\beta}}_r$ be an estimator of $\boldsymbol{\beta}$ based on $\{(\mathbf{x}_{ni}', Y_{ni}^o),\ 1 \le i \le r_n\}$ such that

$$\mathbf{A}_r^{-1}(\hat{\boldsymbol{\beta}}_r - \boldsymbol{\beta}) = \mathbf{T}_r + o_p(1), \qquad \mathbf{T}_r := I_0^{-1}\mathbf{A}_r\mathbf{X}_r'\mathbf{s}_r^o \tag{32}$$

where $s_{ni}^o = -\dot{f}(e_{ni}^o)/f(e_{ni}^o)$, $1 \le i \le r_n$, and $\mathbf{s}_r^o = (s_{ni}^o,\ 1 \le i \le r_n)'$. Define

$$K_1^*(t) := W_1(t) + n^{1/2}\bar{\mathbf{x}}_n'\mathbf{A}_r\cdot\mathbf{T}_r q_0(t), \qquad 0 \le t \le 1.$$

Similar to (28), we obtain, for $s \le t$, that

$$EK_1^*(s)K_1^*(t) = s(1-t) - I_0^{-1}q_0(s)q_0(t)\{\bar{\mathbf{x}}_n'(\mathbf{X}_r'\mathbf{X}_r)^{-1}[n\bar{\mathbf{x}}_n - 2r_n\bar{\mathbf{x}}_r]\}$$

Argue as for Corollary 6.2a.1 to conclude

Corollary 6.2a.2. *Under* (1.1.1), H_0, (NX), (12), (31) *and* (32),

$$\hat{D}_{1r} \xrightarrow[d]{} \sup_{0 \le t \le 1} |B(t)|, \tag{33}$$

where $\hat{D}_{1r}$ *is the* $\hat{D}_1$ *with* $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}_r$. □

Remark 6.2a.4. *Assumptions* (21) *and* (31). To begin with note that if

$$\lim_n n^{-1}(\mathbf{X}_n'\mathbf{X}_n) \text{ exists and is positive definite,} \tag{34}$$

then (21) is equivalent to

$$nm_n^{-1} \longrightarrow 2. \tag{35}$$

If, in addition to (34), one also assumes

$$\lim_n \bar{\mathbf{x}}_n \text{ exists and is finite,} \tag{36}$$

then (31) is equivalent to

$$nr_n^{-1} \longrightarrow 2. \tag{37}$$

There are many designs that satisfy (34) and (36). These include the one way classification, randomized block and the factorial designs, among others.

The choice of m_n and r_n rows is, of course, crucial, and obviously, depends on the design matrix. In the one way classification design with p treatments, n_j observations from the jth treatment, it is recommended to choose the first $m_{nj} = [n_j/2]$ observations from the jth treatment, $1 \le j \le p$, to estimate β. Here $m_n = m_{n1} + \dots + m_{np} = [n/2]$. One chooses $r_{nj} = m_{nj}$, $1 \le j \le p$, $r_n = \Sigma_j r_{nj} = [n/2]$. The choice of m_n and r_n is made similarly in the randomized block design and other similar designs. If one had several replications of a design, where the design matrix satisfies (34) and (36), then one could use the first half of the replications to estimate β and all replications to carry out the test.

Thus, in those cases where designs satisfy (34) and (36), the above construction of the a.d.f. tests is similar to the half sample technique in the one sample problem as found in Rao (1972) or Durbin (1976).

Of course there are designs of interest where (34) and (36) do not hold. An example is $p = 1$, $x_{ni} \equiv i$. Here, $X_n'X_n = O(n^3)$. If one decides to choose the first $m_n(r_n)$ x_i's, then (21) and (31) are equivalent to requiring $(m_n/n)^3 \longrightarrow 1/2$ and $(r_n/n)^2 \longrightarrow 1/2$. Thus, here $\hat{D}_{2m}$ or $\hat{D}_{3m}$ would use 79% of the observations to estimate β while $\hat{D}_{1r}$ would use 71%. On the other hand, if one decides to use the last $m_n(r_n)$ x_i's, then $\hat{D}_2$, $\hat{D}_3$ will use the last 21% observations while $\hat{D}_1$ will use the last 29% observations to estimate β. Of course all of these tests would be based on the entire sample.

In general, to avoid the above kind of problem, one may wish to use, from the practical point of view, some other characteristics of the design matrix in deciding which m_n, r_n rows to choose. One criterion to use may be to choose those $m_n(r_n)$ rows that will approximately maximize (m_n/n) $((r_n/n))$ subject to (21) ((31)). □

Remark 6.2a.5. *Construction of* $\hat{\beta}_m$ and $\hat{\beta}_r$. If F_0 is a d.f. for which the maximum likelihood estimator (m.l.e.) of β has a limiting distribution under (NX) and (12) then one should use this estimator based on r_n (m_n) observations $\{(x_i', Y_i)\}$ for $\hat{D}_1$ ($\hat{D}_2$ or $\hat{D}_3$). For example, if F_0 is the N(0,1) d.f., then the obvious choice for $\hat{\beta}_r$ and $\hat{\beta}_m$ are the least squares estimators:

$$\hat{\beta}_r := (X_r'X_r)^{-1}X_r'Y_r^o; \quad \hat{\beta}_m := (X_mX_m)^{-1}X_m'Y_m^*$$

Of course there are many d.f.'s F_0 that satisfy the above conditions, but for which the computation of m.l.e. is not easy. One way to proceed in such cases is to use one step linear approximation. To make this precise, let $\bar{\beta}_m$ be an estimator of β based on $\{(x_{ni}', Y_{ni}),\ 1 \le i \le m_n\}$ such that

$$A_m^{-1}(\bar{\beta}_m - \beta) = O_p(1). \tag{38}$$

Define

$$(39) \qquad \psi_0(y) := -\dot{f}_0(y)/f_0(y), \qquad\qquad y \in \mathbb{R};$$

$$\bar{s}_{ni} := \psi_0(Y_{ni} - x_{ni}'\bar{\beta}_m), \ 1 \le i \le m_n; \qquad \bar{s}_m := (\bar{s}_{ni}, 1 \le i \le m_n)';$$

$$\hat{\beta}_m := \bar{\beta}_m + I_0 \, A_m A_m \, X_m' s_m;$$

$$V_m^*(y, t) = A_m \sum_{i=1}^{m_n} x_{ni} \, I(Y_{ni} \le y + x_{ni}'t), \qquad\qquad y \in \mathbb{R}, \ t \in \mathbb{R}^p.$$

Then

$$A_m X_m' \bar{s}_m = \int \psi_0(y) \, V_m^*(dy, \bar{\beta}_m).$$

From this and (2.3.37), applied to $\{(x_{ni}', Y_{ni}), \ 1 \le i \le m_n\}$, one readily obtains

Corollary 6.2a.3. *Assume that* (1.1.1) *and* H_0 *hold. In addition, assume that* F_0 *is strictly increasing, satisfies* (12) *and is such that* ψ_0 *is a finite linear combination of nondecreasing bounded functions,* X *and* $\{\bar{\beta}_m\}$ *satisfy* (NX) *and* (38). *Then* $\{\hat{\beta}_m\}$ *of* (39) *satisfies* (24) *for any sequence* $m_n \longrightarrow \infty$, *as* $n \longrightarrow \infty$.

Proof. Clearly,

$$A_m^{-1}(\hat{\beta}_m - \beta) = A_m^{-1}(\bar{\beta}_m - \beta) + I_0^{-1} \, A_m X_m' \bar{s}_m.$$

But, integration by parts and (2.3.37) yield

$$\begin{aligned} A_m X_m' \{\bar{s}_m - s_m\} &= \int \psi_0(y) \{V_m^*(dy, \bar{\beta}) - V_m^*(dy, \beta)\} \\ &= -\int \{V_m^*(y, \bar{\beta}) - V_m^*(y, \beta)\} \, d\psi_0(y) \\ &= -A_m^{-1}(\bar{\beta}_m - \beta) \int f_0(y) d\psi_0(y) + o_p(1) \\ &= -A_m^{-1}(\bar{\beta}_m - \beta) I_0 + o_p(1). \qquad \square \end{aligned}$$

The above result is useful, e.g., when F_0 is logistic, Cauchy or double exponential. In the first case m.l.e. is not easy to compute but F_0 has finite second moment. So take $\bar{\beta}_m$ to be the l.s.e. and then use (39) to obtain the final estimator to be used for testing. In the case of Cauchy, $\bar{\beta}_m$ may be chosen to be an R-estimator.

Clearly, there is an analogue of the above corollary involving $\{\hat{\beta}_r\}$ that would satisfy (31). $\square$

6.2b. Bootstrap Distributions

In this subsection we shall obtain a weak convergence result about a bootstrapped w.e.p.'s and then apply this to yield bootstrap distributions of some of the above tests.

Let (1.1.1) with $e_{ni} \equiv e_i$ and H_0 hold. Let E_0 and P_0 denote the expectation and probability, respectively, under these assumptions. In addition, throughout this section we shall *assume that* ($\mathbf{F_0 1}$), ($\mathbf{F_0 2}$) and (**NX**) hold.

Recall the definition of $\mathbf{W}$, $\hat{\mathbf{W}}$ from (6.2a.1), (6.2a.2). Let $\hat{\boldsymbol{\beta}}$ be an M-estimators of $\boldsymbol{\beta}$ corresponding to a bounded nondecreasing right continuous score function ψ such that

$$(1) \qquad \int \psi \, dF_0 = 0, \qquad \int f_0 \, d\psi > 0.$$

Upon specializing (4.2a.8) to the current setup one readily obtains

$$(2) \qquad \mathbf{A}^{-1}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = -\kappa \, \Sigma_i \, \mathbf{A}\mathbf{x}_{ni} \, \psi(e_i) + o_p(1), \qquad (P_0).$$

where $\kappa := 1/\int f_0 \, d\psi$.

Let the approximating process obtained from (6.2a.5) and (2) be denoted by $\overline{\mathbf{W}}$, i.e.,

$$(3) \qquad \overline{\mathbf{W}}(t) := \Sigma_i \, \mathbf{A}\mathbf{x}_{ni}\{I(e_i \le F_0^{-1}(t)) - t - \kappa \, q_0(t) \, \psi(e_i)\}, \qquad 0 \le t \le 1.$$

Define

$$(4) \qquad \sigma^2 := E_0 \psi^2(e_1),$$

$$g_0(t) := E_0\{I(e_1 \le F_0^{-1}(t)) - t\} \, \psi(e_1)$$

$$= \int I(x \le F_0^{-1}(t)) \, \psi(x) \, dF_0(x), \qquad 0 \le t \le 1,$$

and, for $0 \le t \le u \le 1$,

$$(5) \qquad \rho_0(t, u) := t(1-u) - \kappa \, [q_0(t) g_0(u) + g_0(t) q_0(u)] + \kappa^2 q_0(t) q_0(u) \sigma^2.$$

Note that

$$(6) \qquad \mathcal{C}_0(t, u) := E_0\{\overline{\mathbf{W}}(t)\overline{\mathbf{W}}(u)'\} = \rho_0(t, u) \mathbf{I}_{p \times p}, \qquad 0 \le t \le u \le 1.$$

Let $\mathcal{G}_0 := (\mathcal{G}_{01}, \ldots, \mathcal{G}_{0p})'$ be a p-vector of independent Gaussian processes each having the covariance function ρ_0. Thus, $E\mathcal{G}_0(t)\mathcal{G}_0(u)' \equiv \mathcal{C}_0(t, u)$. Since ρ_0 is continuous, $\mathcal{G}_0 \in \{\mathbb{C}[0, 1]\}^p$. Moreover, from Corollary 2.2a.1 applied p time, j^{th} time to the entities $X_{ni} \equiv e_i$, $F_{ni} \equiv F_0$ and $d_{ni} \equiv (i,j)^{th}$

entry of **AX**, $1 \le j \le p$, $1 \le i \le n$, and from the uniform continuity of q_0 it readily follows that

(7) $$\overline{\mathbf{W}} \Rightarrow \mathcal{G}_0 \text{ in } [\{\mathbb{D}[0,1]\}^p, \mathcal{d}].$$

Now, let $\hat{f}_n$ be a density estimator based on $\{\hat{e}_{ni} := Y_{ni} - \mathbf{x}_{ni}'\hat{\boldsymbol{\beta}};\ 1 \le i \le n\}$ and $\hat{F}_n$ be the corresponding d.f.. Let $\{e_{ni}^*;\ 1 \le i \le n\}$ represent i.i.d. $\hat{F}_n$ r.v.'s, i.e., $\{e_{ni}^*;\ 1 \le i \le n\}$ is a random sample from the population $\hat{F}_n$. Because $\hat{F}_n$ is continuous, the resampling procedures based on it are usually called *smooth* bootstrap procedures. Let

(8) $$Y_{ni}^* := \mathbf{x}_{ni}'\hat{\boldsymbol{\beta}} + e_{ni}^*, \qquad 1 \le i \le n.$$

Define the bootstrap estimator $\boldsymbol{\beta}^*$ to be a solution $\mathbf{s} \in \mathbb{R}^p$ of the equation

(9) $$\Sigma_i\, \mathbf{A}\mathbf{x}_{ni}\, \{\psi(Y_{ni}^* - \mathbf{x}_{ni}'\mathbf{s}) - \hat{E}_n \psi(e_{n1}^*)\} = 0.$$

where $\hat{E}_n$ is the expectation under $\hat{F}_n$. Let $\hat{P}_n$ denote the the bootstrap probability under $\hat{F}_n$. Finally, define

(10) $$\mathcal{S}^*(t, \mathbf{u}) := \Sigma_i\, \mathbf{A}\mathbf{x}_{ni}\, I(e_{ni}^* \le \hat{F}_n^{-1}(t) + \mathbf{x}_{ni}'\mathbf{A}\mathbf{u}), \qquad 0 \le t \le 1,\ \mathbf{u} \in \mathbb{R}^p,$$

and the vector of bootstrap w.e.p.'s

(11) $$\hat{\mathbf{W}}^*(t) := \Sigma_i\, \mathbf{A}\mathbf{x}_{ni}\, \{I(Y_{ni}^* - \mathbf{x}_{ni}'\boldsymbol{\beta}^* \le \hat{F}_n^{-1}(t)) - t\}, \qquad 0 \le t \le 1.$$

We also need

(12) $$\mathbf{W}^*(t) := \Sigma_i\, \mathbf{A}\mathbf{x}_{ni}\{I(e_{ni}^* \le \hat{F}_n^{-1}(t)) - t\}, \qquad 0 \le t \le 1.$$

Our goal is to show that $\hat{\mathbf{W}}^*$ converges weakly to $\mathcal{G}_0$ in $[\{\mathbb{D}[0,1]\}^p, \mathcal{d}]$, a.s.. Here a.s. refers to almost all error sequences $\{e_i;\ i \ge 1\}$. We in fact have the following

Theorem 6.2b.1. *In addition to* (1.1.1), H_0, $(\mathbf{F}_0\mathbf{1})$, $(\mathbf{F}_0\mathbf{2})$, (**NX**) *and* (1), *assume that* ψ *is a bounded nondecreasing right continuous score function and that the following hold.*

(13) *For almost all error sequences* $\{e_i;\ i \ge 1\}$, $\hat{f}_n(x) > 0$ *for almost all* $x \in \mathbb{R}$, $n \ge 1$.

(14) $$\|\hat{f}_n - f_0\|_\infty \longrightarrow 0, \text{ a.s., } (P_0).$$

Then, $\forall\ 0 < B < \infty$,

(15) $$\sup \|\mathcal{S}^*(t, \mathbf{u}) - \mathcal{S}^*(t, \mathbf{0}) - \mathbf{u}\hat{f}_n(\hat{F}_n^{-1}(t))\| = o_p(1),\ (\hat{P}_n), \text{ a.s.},$$

where the supremum is over $0 \le t \le 1$, $\|\mathbf{u}\| \le B$.

Moreover, for almost all error sequences $\{e_i;\ i \ge 1\}$,

$$(16)\qquad \mathbf{A}^{-1}(\beta^* - \hat{\beta}) = -\hat{\kappa}_n \Sigma_i \mathbf{A}\mathbf{x}_{ni}\{\psi(e^*_{ni}) - \hat{E}_n\psi(e^*_{n1})\} + o_p(1), \quad (\hat{P}_n),$$

and

$$(17)\qquad \hat{\mathbf{W}}^* \Rightarrow \mathcal{G}_0 \ \text{ in } [\{\mathbb{D}[0,1]\}^p, d],$$

where $\hat{\kappa}_n := 1 / \int \hat{f}_n \, d\psi$.

Proof. Fix an error sequences $\{e_i;\ i \ge 1\}$ for which

$$(14^*)\qquad \hat{f}_n(x) > 0, \text{ for almost all } x \in \mathbb{R}, \text{ and } \|\hat{f}_n - f_0\|_\infty \longrightarrow 0.$$

The following arguments are carried out conditional on this sequence.

Observe that $\mathcal{S}^*(t, \mathbf{u})$ is a p-vector of w.e.p.'s $S_d(t, \mathbf{u})$ of (2.3.1) whose j^{th} component has various underlying entities as follows:

$$(18)\qquad X_{ni} = e^*_{ni},\ F_{ni} = \hat{F}_n,\ \mathbf{c}_{ni} = \mathbf{A}\mathbf{x}_{ni},\ d_{ni} = \mathbf{a}'_{(j)}\mathbf{x}_{ni},\ 1 \le i \le n$$

where, as usual, $\mathbf{a}_{(j)} = j^{th}$ column of $\mathbf{A}$, $1 \le j \le p$.

Thus, (15) follows from p applications of Theorem 2.3.1, j^{th} time applied to the above entities, provided we ensure the validity of the assumptions of that theorem. But, f_0 uniformly continuous and (14) readily imply that $\{\hat{f}_n, n \ge 1\}$ satisfies (2.3.3a,b). In view of (2.3.33), (2.3.34) and (NX), it follows that all other assumptions of Theorem 2.3.1 are satisfied. Hence, (15) follows from (2.3.6). In view of (13) we also obtain, from (2.3.7),

$$(19)\qquad \sup \|\mathcal{S}^{0*}(x, \mathbf{u}) - \mathcal{S}^{0*}(x, \mathbf{0}) - \mathbf{u}\hat{f}_n(x)\| = o_p(1), \quad (\hat{P}_n),$$

where $\mathcal{S}^{0*}(x, \mathbf{u}) \equiv \mathcal{S}^*(\hat{F}_n(x), \mathbf{u})$ and where the supremum is over $x \in \mathbb{R}$, $\|\mathbf{u}\| \le B$. Now, (16) follows from (19) in precisely the same fashion as does (4.2a.8) from (2.3.7).

From (11), (15), (16) and (31) below, we readily obtain that, under $\hat{P}_n$,

$$(20)\qquad \hat{\mathbf{W}}^*(t) = \Sigma_i \mathbf{A}\mathbf{x}_{ni}\{I(e^*_{ni} \le \hat{F}_n^{-1}(t)) - t - \hat{\kappa}_n \hat{q}_n(t)[\psi(e^*_{ni}) - \hat{E}_n\psi(e^*_{n1})]\} + o_p(1),$$

where $\hat{q}_n := \hat{f}_n(\hat{F}_n^{-1})$.

In analogy to (4) and (5), let $\hat{g}_n$, $\hat{\rho}_n$ stand for g_0, ρ_0 after F_0 is replaced by $\hat{F}_n$ in these entities. Thus

$$(21)\qquad \hat{g}_n(t) := \hat{E}_n\{I(e^*_{n1} \le \hat{F}_n^{-1}(t)) - t\}\ \psi(e^*_{n1}) = \int I(x \le \hat{F}_n^{-1}(t))\ \psi(x)\, d\hat{F}_n(x), \qquad 0 \le t \le 1,$$

and, for $0 \le t \le u \le 1$,

(22) $$\hat{\rho}_n(t, u) := t(1-u) - \hat{\kappa}_n[\hat{q}_n(t)\hat{g}_n(u)+\hat{g}_n(t)\hat{q}_n(u)] + \hat{\kappa}_n^2\hat{q}_n(t)\hat{q}_n(u)\hat{\sigma}_n^2.$$

where $\hat{\sigma}_n^2 := \hat{E}_n[\psi(e_{n1}^*) - \hat{E}_n\psi(e_{n1}^*)]^2$.

Let $\tilde{W}^*(t)$ denote the leading r.v. in the r.h.s. of (20). Observe that,

(23) $$\tilde{\mathcal{C}}_n(t, u) := \hat{E}_n\{\tilde{W}^*(t)\tilde{W}^*(u)'\} = \hat{\rho}_n(t, u)\, I_{pxp}, \qquad 0 \le t \le u \le 1.$$

(24) **Claim:** $\hat{\rho}_n(t, u) \longrightarrow \rho_0(t, u), \quad \forall\ 0 \le t \le u \le 1.$

To prove (24), note that (14*) and Scheffé's Theorem (Lehmann, 1986, p573) imply that for the given error sequence $\{e_i;\ i \ge 1\}$,

(25) $$\delta_n := \|\hat{F}_n - F_0\|_\infty \longrightarrow 0,$$

which, together with the continuity of $\hat{F}_n$, yields

(26) $$\sup_{0\le t\le 1} |F_0(\hat{F}_n^{-1}(t)) - t| \longrightarrow 0.$$

Also, observe that

$$\sup_{0\le t\le 1} |\hat{f}_n(\hat{F}_n^{-1}(t)) - f_0(\hat{F}_n^{-1}(t))| \le \|\hat{f}_n - f_0\|_\infty \longrightarrow 0,$$

by (14*), and that,

$$|f_0(\hat{F}_n^{-1}(t)) - f_0(F_0^{-1}(t))| \equiv |q_0(F_0(\hat{F}_n^{-1}(t))) - q_0(t)|, \qquad \forall\ 0 \le t \le 1.$$

Hence, by (26) and the uniform continuity of q_0, which is implied by ($\mathbf{F}_0$**1**),

(27) $$\sup_{0\le t\le 1}|\hat{q}_n(t) - q_0(t)| \longrightarrow 0.$$

Next, let $g_n(t) = \int I(\hat{F}_n(x) \le t)\psi(x)f_0(x)\, dx, \quad 0 \le t \le 1.$ Upon rewriting $\hat{g}_n(t) = \int I(\hat{F}_n(x) \le t)\psi(x)\hat{f}_n(x)\, dx$, from (14*), Scheffé's Theorem (Lehmann: 1986, p 573) and the boundedness of ψ, we readily obtain that

$$\sup_{0\le t\le 1}|\hat{g}_n(t) - g_n(t)| \le \int |\hat{f}_n(x) - f_0(x)|\, dx \longrightarrow 0.$$

But, the inequality $F_0(x) - \delta_n \le \hat{F}_n(x) \le F_0(x) + \delta_n$ for all x, implies that

$$\begin{aligned}|g_n(t) - g_0(t)| &\le \|\psi\|_\infty \int I(F_0(x) - \delta_n \le t \le F_0(x) + \delta_n)\, dF_0(x),\\ &\le \|\psi\|_\infty\, 2\delta_n, \qquad \forall\ 0 \le t \le 1.\end{aligned}$$

Hence, by (25),

(28) $$\sup_{0\le t\le 1}|\hat{g}_n(t) - g_0(t)| \longrightarrow 0.$$

Again by the boundedness of ψ, (14*) and (25), one readily concludes that

(29) $$\hat{\kappa}_n \longrightarrow \kappa, \quad \hat{\sigma}_n^2 \longrightarrow \sigma^2.$$

Claim (24) now readily follows from (27) – (29).

Now recall (12) and rewrite $\tilde{\mathbf{W}}^*$ as

(30) $$\tilde{\mathbf{W}}^*(t) = \mathbf{W}^*(t) - \hat{\kappa}_n \hat{q}_n(t) \Sigma_i \mathbf{A}\mathbf{x}_{ni} [\psi(e^*_{ni}) - \hat{E}_n\psi(e^*_{n1})].$$

Observe that because

$$\hat{E}_n\|\Sigma_i \mathbf{A}\mathbf{x}_{ni} [\psi(e^*_{ni}) - \hat{E}_n\psi(e^*_{n1})]\|^2 = p\, \hat{\sigma}_n^2,$$

by (29) and the Markov inequality it follow that

(31) $$\|\Sigma_i \mathbf{A}\mathbf{x}_{ni} [\psi(e^*_{ni}) - \hat{E}_n\psi(e^*_{n1})]\| = O_p(1), \ (\hat{P}_n).$$

Apply Corollary 2.2a.1 p times, j^{th} time to the entities given at (18), to conclude that

$$\lim_{\eta\to 0} \limsup_n \hat{P}_n\Big(\sup_{|t-s|<\eta} |\mathbf{W}^*(t) - \mathbf{W}^*(s)| > \eta\Big) = 0.$$

This together with (31), (30), (27) and the uniform continuity of F_0 implies that the sequence of processes $\{\tilde{\mathbf{W}}^*\}$ is tight in the uniform metric $\mathscr{d}$ and all its subsequential limits must be in $\{\mathbb{C}[0, 1]\}^p$. Now, (17) follows from this, Claim (24), (20), (13), (14) and (6). □

Remark 6.2b.1. One of the main consequences of (17) is that one can use the bootstrap analogue of $\hat{D}_3$, v.i.z., $\hat{D}_3^* := \sup\{\|\hat{\mathbf{W}}^*(t)\|, 0\le t\le 1\}$ to carry out the test H_0. Thus an approximation to the the null distribution of $\hat{D}_3$ is obtained by the distribution of $\hat{D}_3^*$ under $\hat{P}_n$. In practice it means to obtain repeated random samples of size n from $\hat{F}_n$, compute the frequency distribution of $\hat{D}_3^*$ from these samples and use that to approximate the null distribution of $\hat{D}_3$. At least asymptotically this converges to the right distribution. Obviously the smooth bootstrap distributions for $\hat{D}_1$, $\hat{D}_2$ can be obtained similarly.

Reader might have realized that the conclusion (17) is true for any sequence of estimators $\{\hat{\beta}\}$, $\{\beta^*\}$ satisfying (2) and (16). □

6.3. L_2-DISTANCE TESTS

Let K_1^o and K_2^o, respectively, stand for the K_1 and $K_{\mathbf{X}}$ of (5.2.5) and (5.2.7) after the d.f.'s $\{H_{ni}\}$ there are replaced by F_0. Thus, for $G \in \mathcal{DI}(\mathbb{R})$,

$$(1) \qquad K_1^o(\mathbf{t}) := \int \{W_1^o(y, \mathbf{t})\}^2 dG(y),$$

$$K_2^o(\mathbf{t}) := \int \|\mathbf{W}^o(y, \mathbf{t})\|^2 dG(y), \qquad \mathbf{t} \in \mathbb{R}^p,$$

where $\mathbf{W}^o$ is as in (6.1.3) and

$$(2) \qquad W_1^o(y, \mathbf{t}) := n^{1/2}[H_n(y, \mathbf{t}) - F_0(y)], \qquad y \in \mathbb{R},\ \mathbf{t} \in \mathbb{R}^p.$$

Let $\hat{\boldsymbol{\beta}}$ be an estimator of $\boldsymbol{\beta}$ and define the four test statistics

$$(3) \qquad K_j^* := \inf\{K_j(\mathbf{t});\ \mathbf{t} \in \mathbb{R}^p\}, \qquad \hat{K}_j := K_j(\hat{\boldsymbol{\beta}}), \qquad j = 1, 2.$$

The large values of these statistics are significant for testing H_0.

We shall first discuss the a.n.d.'s of K_j^*, $j = 1, 2$. Let $W_1^o(\cdot)$, $\mathbf{W}^o(\cdot)$ stand for $W_1^o(\cdot, \boldsymbol{\beta})$ and $\mathbf{W}^o(\cdot, \boldsymbol{\beta})$.

Theorem 6.3.1. *Assume that* (1.1.1), H_0, (NX), (5.5.68) – (5.5.70) *with* $F \equiv F_0$ *hold.*

(a) *If, in addition,* (5.6a.10) *and* (5.6a.11) *hold, then*

$$(4) \qquad K_1^* = \int \Big\{W_1^o(y) - f_0(y)\, \frac{\int W_1^o f_0\, dG}{\int f_0^2 dG}\Big\}^2 dG + o_p(1).$$

(b) *Under no additional assumptions,*

$$(5) \qquad K_2^* = \int \Big\|\mathbf{W}^o(y) - f_0(y)\, \frac{\int \mathbf{W}^o f_0\, dG}{\int f_0^2 dG}\Big\|^2 dG + o_p(1).$$

Proof. Apply Theorems 5.5.1 and 5.5.3 twice, once with $\mathbf{D} = n^{-1/2}(1, 0, \ldots, 0]$ and once with $\mathbf{D} = \mathbf{XA}$, and the rest of the entities as follows:

$$(6) \qquad Y_{ni} \equiv e_{ni}, \qquad H_{ni} \equiv F_0 \equiv F_{ni}, \qquad G_n \equiv G.$$

The theorem then follows from (5.5.28), (5.6a.5), (5.6a.12) and some algebra. See also Claim 5.5.2. □

Remark 6.3.1. Perhaps it is worthwhile repeating that (5) holds without any extra conditions on the design matrix $\mathbf{X}$. Thus, at least in this

sense, K_2^* is a more natural statistic to use than K_1^* for testing H_0.

A consequence of (4) is that even if $\hat{\beta}_1$ of (5.2.4) is asymptotically non–unique, K_1^* asymptotically behaves like a unique sequence of r.v.'s. Moreover, unlike the $\hat{D}_1$–statistic, the asymptotic null distribution of K_1^* does not depend on the design matrix among all those designs that satisfy the given conditions.

The assumptions (5.6a.10) and (5.6a.11) are restrictive. For example, in the case $p = 1$, (5.6a.10) translates to requiring that either $x_{i1} \geq 0$ for all i or $x_{i1} \leq 0$ for all i. The assumption (5.6a.11) says that $\bar{\mathbf{x}} \neq \mathbf{0}$ or can not converge to $\mathbf{0}$. Compare this with the fact that if $\bar{\mathbf{x}} \approx 0$ then the asymptotic distribution of $\hat{D}_1$ does not depend on the preliminary estimator $\hat{\beta}$. □

Next, we need a result that will be useful in deriving the limiting distributions of certain quadratic forms involving w.e.p.'s. To that effect, let $L_2^p(\mathbb{R}, G)$ be the equivalence classes of measurable functions h: $\mathbb{R}$ to $\mathbb{R}^p$ such that $|h|_G^2 := \int \|h\|^2 dG < \infty$. The equivalence classes are defined in terms of the norm $|\cdot|_G^2$. In the following lemma, $\{\mathbf{a}_i;\ i \geq 1\}$ is a fixed orthonormal basis in $L_2^p(\mathbb{R}, G)$.

Lemma 6.3.1. *Let* $\{\mathbf{Z}_n,\ n \geq 1\}$ *be a sequence of p–vector stochastic processes with* $E\mathbf{Z}_n = \mathbf{0}$, $\mathrm{Cov}(\mathbf{Z}_n(x),\ \mathbf{Z}_n(y)) := \mathbf{K}_n(x, y) = ((K_{nij}(x, y)))$, $1 \leq i, j \leq p$, $x, y \in \mathbb{R}$. *In addition, assume the following:*

There is a covariance matrix function $\mathbf{K}(x, y) = ((K_{ij}(x, y)))$, *and a p–vector mean zero covariance–*$\mathbf{K}$ *Gaussian process* $\mathbf{Z}$ *such that*

(i) (a) $\quad \sum_{j=1}^{p} \int K_{njj}(x, x)dG(x) < \infty,\ n \geq 1.$ (b) $\sum_{j=1}^{p} \int K_{jj}(x, x)dG(x) < \infty.$

(ii) $\quad \sum_{j=1}^{p} \int K_{njj}(x, x)dG(x) \longrightarrow \sum_{j=1}^{p} \int K_{jj}(x, x)dG(x).$

(iii) $\quad$ *For every* $m \geq 1$,

$$\left(\int \mathbf{Z}_n'\mathbf{a}_1 dG, \ldots, \int \mathbf{Z}_n'\mathbf{a}_m dG\right) \underset{d}{\longrightarrow} \left(\int \mathbf{Z}'\mathbf{a}_1 dG, \ldots, \int \mathbf{Z}'\mathbf{a}_m dG\right);$$

(iv) $\quad$ *For each* $i \geq 1$,

$$E\left(\int \mathbf{Z}_n'\mathbf{a}_i dG\right)^2 \longrightarrow E\left(\int \mathbf{Z}'\mathbf{a}_i dG\right)^2.$$

Then, $\mathbf{Z}_n$, $\mathbf{Z}$ *belong to* $L_2^p(\mathbb{R}, G)$, *and*

$$\mathbf{Z}_n \Rightarrow \mathbf{Z} \text{ in } L_2^p(\mathbb{R}, G). \tag{7}$$

Proof: In view of Theorem VI.2.2 of Parthasarthy (1967) and in view of (iii), it suffices to show that for any $\epsilon > 0$, there is an $N\ (= N_\epsilon)$ such that

$$\sup_n E \sum_{i \geq N} \left(\int Z_n' \mathbf{a}_i dG\right)^2 \leq \epsilon. \tag{8}$$

Because of the properties of $\{\mathbf{a}_i\}$, Fubini and (i),

$$\sum_{j=1}^{p} \int K_{njj}(x, x) dG(x) = E|Z_n|_G^2 = \sum_{i \geq 1} E\left(\int Z_n' \mathbf{a}_i dG\right)^2, \tag{9}$$

$$\sum_{j=1}^{p} \int K_{jj}(x, x) dG(x) = E|Z|_G^2 = \sum_{i \geq 1} E\left(\int Z' \mathbf{a}_i dG\right)^2. \tag{10}$$

Thus, to prove (8), it suffices to exhibit an N such that

$$\sup_n \sum_{i \geq N} E\left(\int Z_n' \mathbf{a}_i dG\right)^2 \leq \epsilon. \tag{11}$$

By (ii), (9) and (10), there exists $N_{1\epsilon}$ such that

$$\sum_{i \geq 1} E\left(\int Z_n' \mathbf{a}_i dG\right)^2 \leq \sum_{i \geq 1} E\left(\int Z' \mathbf{a}_i dG\right)^2 + \epsilon/3, \qquad n \geq N_{1\epsilon}. \tag{12}$$

By (i)(b) and (10), there exists $N(= N_\epsilon)$ such that

$$\sum_{i \geq N} E\left(\int Z' \mathbf{a}_i dG\right)^2 \leq \epsilon/3. \tag{13}$$

By (iv), there exists $N_{2\epsilon}$ such that

$$\sum_{i < N} E\left(\int Z' \mathbf{a}_i dG\right)^2 \leq \sum_{i < N} E\left(\int Z_n' \mathbf{a}_i dG\right)^2 + \epsilon/3, \qquad n \geq N_{2\epsilon}. \tag{14}$$

Therefore, from (12) – (14), with $N = N_\epsilon := N_{1\epsilon} \vee N_{2\epsilon}$,

$$\begin{aligned} &\sup_{n \geq N} \sum_{i < N} E\left(\int Z_n' \mathbf{a}_i dG\right)^2 \\ &\qquad \leq \sup_{n \geq N} \left[\sum_{i \geq 1} E\left(\int Z' \mathbf{a}_i dG\right)^2 - \sum_{i < N} \left(\int Z_n' \mathbf{a}_i dG\right)^2\right] + \epsilon/3 \leq \epsilon. \end{aligned}$$

Use (i)(a) to take care of the case $n < N_\epsilon$. This proves the result. □

Remark 6.3.2. Millar (1981) contains a special case of the above lemma where $p = 1$, Z_n is the standardized ordinary e.p. and Z is the Brownian

bridge. The above lemma is an extension of Millar's result to cover more general processes like the w.e.p.'s under general independent setting. In applications of the above lemma, one may choose $\{\mathbf{a}_i\}$ to be such that the support S_i of $\mathbf{a}_i$ has $G(S_i) < \infty$, $i \geq 1$ and such that $\{\mathbf{a}_i\}$ are bounded. □

Corollary 6.3.1. (a) *Under the conditions of Theorem* 6.3.1(a),

$$K_1^* \underset{d}{\longrightarrow} \int \Big\{ \mathcal{B}(F_0) - f_0 \cdot \frac{\int \mathcal{B}(F_0) f_0 \, dG}{\int f_0^2 dG} \Big\}^2 dG =: \overline{G}_1, \qquad \text{(say)}. \tag{15}$$

(b) *Under the conditions of Theorem* 6.3.1(b),

$$K_2^* \underset{d}{\longrightarrow} \int \Big\| \mathbf{\mathcal{B}}(F_0) - f_0 \cdot \frac{\int \mathbf{\mathcal{B}}(F_0) f_0 \, dG}{\int f_0^2 dG} \Big\|^2 =: \overline{G}_2, \qquad \text{(say)}. \tag{16}$$

Here $\mathcal{B}$, $\mathbf{\mathcal{B}}$ *are is as in* (6.2a.7), (6.2a.8).

Proof: (b) Apply Lemma 6.3.1, with $\mathbf{a}_i$ as in the Remark 6.3.2 above, to

$$\mathbf{Z}_n = \mathbf{W}^o - \frac{\int \mathbf{W}^o f_0 \, dG}{\int f_0^2 dG} \cdot f_0, \qquad \mathbf{Z} = \mathbf{\mathcal{B}}(F_0) - \frac{\int \mathbf{\mathcal{B}}(F_0) f_0 dG}{\int f_0^2 dG} \cdot f_0.$$

Direct calculations show that $E\mathbf{Z}_n = \mathbf{0} = E\mathbf{Z}$, and $\forall\ x, y \in \mathbb{R}$,

$$K_n(x, y) := E\mathbf{Z}_n(x)\mathbf{Z}_n'(y) = I_{p\times p}\, \ell(x, y) = K(x, y) =: E\mathbf{Z}(x)\mathbf{Z}'(y),$$

where, for $x, y \in \mathbb{R}$,

$$\ell(x, y) := k(x, y) - a^{-1} f_0(y) \int k(x, s)\, d\psi(s) - a^{-1} f_0(y) \int k(y, s)\, d\psi(s) + \\ + a^{-2} \int\int k(s, t)\, d\psi(s) d\psi(t),$$

$$k(x, y) := F_0(x \wedge y) - F_0(x) F_0(y), \quad \psi(x) = \int_{-\infty}^{x} f_0 dG, \quad a = \psi(\infty).$$

Therefore, (5.5.68), (5.5.69) imply (i), (ii) and (iv). To prove (iii), let $\lambda_1, \ldots, \lambda_m$ be real numbers. Then,

$$\sum_{j=1}^{m} \lambda_j \int \mathbf{Z}_n' \mathbf{a}_j dG = \int \mathbf{W}^{o\prime} \mathbf{b}\, dG - \frac{\int \mathbf{W}^{o\prime} d\psi}{\int f_0 d\psi} \cdot \int \mathbf{b}\, d\psi =: h(\mathbf{W}^o), \qquad \text{(say)},$$

where $\mathbf{b} := \sum_{j=1}^{m} \lambda_j \mathbf{a}_j$. Because ψ and $\mathbf{b}dG$ are finite measures, $h(\mathbf{W}^o)$ is a uniformly continuous function of $\mathbf{W}^o$. Thus by Lemma 6.2a.2 and Theorem

5.1 of Billingsley (1968), $h(\mathbf{W}^o) \xrightarrow[d]{} h(B(F_0))$, under H_0 and (NX). This then verifies all conditions of Lemma 6.3.1. Hence $\mathbf{Z}_n \Rightarrow \mathbf{Z}$ in $L_2^p(\mathbb{R}, G)$. In particular $\int \|Z_n\|^2 dG \xrightarrow[d]{} \int \|Z\|^2 dG$. This and (5) proves (16). The proof of (15) is similar. □

Remark 6.3.3. The r.v. $\overline{G}_1$ can be rewritten as

$$\overline{G}_1 = \int B^2(F_0)dG - \frac{\{\int B(F_0) f_0 dG\}^2}{\int f_0^2 dG}$$

Recall that $\overline{G}_1$ is the same as the limiting r.v. obtained in the one sample location model. Its distribution for various G and F_0 has been theoretically studied by Martynov (1975). Boos (1981) has tabulated some critical values of $\overline{G}_1$ when $dG = \{F_0(1 - F_0)\}^{-1} dF_0$ and F_0 = Logistic. From Anderson–Darling or Boos one obtains that in this case

$$\overline{G}_1 = \int_0^1 B^2(t)(t(1-t))^{-1}dt - 6\Big(\int_0^1 B(t)dt\Big)^2 = \sum_{j\geq 2} N_j^2/j(j+1)$$

where $\{N_j\}$ are i.i.d. $N(0, 1)$ r.v.'s. From Boos (Table 3), one obtains the following

Table II

α	.005	.01	.025	.05
t_α	1.710	1.505	1.240	1.046

In Table II, t_α is such that $P(\overline{G}_1 > t_\alpha) = \alpha$. For some other tables see Stephens (1979).

The r.v. $\overline{G}_2$ can be rewritten as

$$\begin{aligned}\overline{G}_2 &:= \int \|B(F_0)\|^2 dG - \frac{\|\int B(F_0) f_0 dG\|^2}{\int f_0^2 dG} \\ &= \sum_{j=1}^{p} \Big[\int B_j^2(F_0)dG - \frac{(\int B_j(F_0) f_0 dG)^2}{\int f_0^2 dG}\Big],\end{aligned}$$

which is a sum of p independent r.v.'s identically distributed as $\overline{G}_1$. The distribution of such r.v.'s does not seem to have been studied yet. Until the distribution of $\overline{G}_2$ is tabulated one could use the independence of the

summands in $\overline{G}_2$ and the bounds between the sum and the maximum to obtain a crude approximation to the significance level.

For p = 1, the a.n.d. of K_1^* and K_2^* is the same but the conditions under which the results for K_1^* hold are stronger than those for K_2^*. □

The next result gives an approximation for $\hat{K}_j$, j = 1, 2. It also follows from Theorem 5.5.1 in a fashion similar to the previous theorem, and hence no details are given.

Theorem 6.3.2. *Assume that* (1.1.1), H_0, (NX), (5.5.68) – (5.5.70) *with* $F \equiv F_0$ *and* (6.2a.3) *hold. Then,*

$$\hat{K}_1 = \int [W_1^o(y) + n^{1/2}\,\bar{x}A\cdot A^{-1}(\hat{\beta}-\beta)f_0(y)]^2 dG(y) + o_p(1). \tag{17}$$

$$\hat{K}_2 = \int \|W^o(y) + A^{-1}(\hat{\beta}-\beta)f_0(y)\|^2 dG(y) + o_p(1). \quad □$$

From this we can obtain the asymptotic null distribution of these statistics when $\hat{\beta}$ is estimated efficiently for the large samples as follows. Recall the definition of $\{s_i\}$ from (6.2a.13) and let

$$\gamma_i(y) := I(e_i \le y) - F_0(y) + n\bar{x}'(X'X)^{-1}x_i\, s_i \mathcal{I}_0^{-1} f_0(y),$$

$$\alpha_i(y) := I(e_i \le y) - F_0(y) + s_i \mathcal{I}_0^{-1} f_0(y), \qquad 1 \le i \le n, \qquad y \in \mathbb{R},$$

$$\alpha = (\alpha_1, \dots, \alpha_n)', \qquad \gamma = (\gamma_1, \dots, \gamma_n)'.$$

Also, define

$$Z_{n1}(y) := W_1^o(y) + n^{1/2}\bar{x}'AAX's\mathcal{I}_0^{-1}f_0(y) = n^{-1/2}\sum_{i=1}^n \gamma_i(y) \tag{19}$$

$$Z_{n2}(y) := W^o(y) + AX's\mathcal{I}_0^{-1}f_0(y) = AX'\alpha(y), \qquad y \in \mathbb{R}.$$

From Theorem 6.3.2 we readily obtain the

Corollary 6.3.2. *Assume that* (1.1.1), H_0, (NX), (5.5.68) – (5.5.70) *with* $F \equiv F_0$, (6.2a.12) *and* (6.2a.14) *hold. Then,*

$$\hat{K}_1 = \int Z_{n1}^2 dG + o_p(1). \tag{20}$$

$$\hat{K}_2 = \int \|Z_{n2}\|^2 dG + o_p(1). \tag{21}$$ □

Next, observe that for $y \le z$,

$$\begin{aligned} K_{n1}(y, z) &:= \mathrm{Cov}(Z_{n1}(y), Z_{n1}(z)) \\ &= F_0(y)(1-F_0(z)) - n\bar{x}'(X'X)^{-1}\bar{x}\,\frac{f_0(y)f_0(z)}{I_0} =: \ell_{n1}(y, z), \\ K_{n2}(y, z) &:= EZ_{n2}(y)Z_{n2}'(z) \\ &= \{F_0(y)(1 - F_0(z)) - \frac{f_0(y)f_0(z)}{I_0}\}I_{p\times p} =: r_0(y, z), \quad \text{say.} \end{aligned}$$

Now apply Lemma 6.3.1 and argue just as in the proof of Corollary 6.3.1 to conclude

Corollary 6.3.3. (a). *In addition to the conditions of Corollary* 6.3.2, *assume that*

$$n\bar{x}'(X'X)^{-1}\bar{x} \longrightarrow c, \quad |c| < \infty. \tag{22}$$

Then,

$$\hat{K}_1 \xrightarrow[d]{} \int Z_1^2(y)dG(y) \tag{23}$$

where Z_1 *is a Gaussian process in* $L_2(\mathbb{R}, G)$ *with the covariance function*

$$K_1(x, y) := F_0(x)(1 - F_0(y)) - cf_0(x)f_0(y)\, I_0^{-1}, \qquad x \le y. \tag{24}$$

(b) *Under the conditions of Corollary* 6.3.2,

$$\hat{K}_2 \xrightarrow[d]{} \int \| Y_0\|^2 dG \tag{25}$$

where Y_0 *is a vector of* p *independent Gaussian processes in* $L_2^p(\mathbb{R}, G)$ *with the covariance matrix* $r_0 \cdot I_{p\times p}$. □

Remark 6.3.4. Again, observe that the test statistic $\hat{K}_1$ based on the ordinary empirical of the residuals has an a.n.d. which is design dependent whereas the a.n.d. of the test based on the weighted empiricals $\hat{K}_2$ is design free. In fact, for $p = 1$, the limiting r.v. in (25) is the same as the one that appears in the one sample location model. For $G = F_0 = N(0, 1)$ d.f., Martynov (1976) has tabulated the distribution of this r.v.. Stephens (1976) has also tabulated the distribution of this r.v. for $G = F_0$, $dG = dG_0 = \{F_0(1 - F_0)\}^{-1}dF_0$, and for $F_0 = N(0, 1)$. For $G = F_0$, $F_0 = N(0, 1)$ d.f., Stephens and Martynov's tables generally agree up to the two decimal places, though occasionally there is an agreement up to three decimal places. In any case, for $p = 1$, one could use these tables to implement the test based on $\hat{K}_2$, at least asymptotically, whereas the test based on $\hat{K}_1$, being design dependent, can not be readily implemented. For the sake of convenience we reproduce some of the Stephens (1976, 1979) tables below.

Table III

$F_0 = N(0, 1)$

$\hat{K}_2$ \ α	0.10	.025	.05	.10
$\hat{K}_2(F_0)$	.237	.196	.165	.135
$\hat{K}_2(G_0)$	1.541	1.281	1.088	.897

In Table III, $\hat{K}_2(G)$ stands for the $\hat{K}_2$ with G being the integrating measure. $\hat{K}_2(G_0)$ is the $\hat{K}_2$ with the Anderson–Darling weights. Table III is, of course, useful only when $p = 1$. □

As far as the *asymptotic power* of the above L_2-tests is concerned, it is apparent that Theorems 5.5.1, 5.5.3 and Lemma 6.3.1 can be used to deduce the asymptotic power of these tests against fairly general alternatives. Here we shall discuss the asymptotic behavior of only K_j^*, $j = 1, 2$ under the heteroscedastic gross errors alternatives. More precisely, suppose that

$$F_{ni} = (1 - \delta_{ni})F_0 + \delta_{ni}F_1, \quad 0 \le \delta_{ni} \le 1, \quad \max_i \delta_{ni} \longrightarrow 0, \tag{26}$$

F_1 a fixed d.f. Let

$$m_1 := n^{-1/2} \Sigma_i\, \delta_{ni}(F_1 - F_0), \qquad \mathbf{m}_2 := \Sigma_i\, \mathbf{A}\mathbf{x}_{ni}\delta_{ni}(F_1 - F_0).$$

Lemma 6.3.2. *Let* (1.1.1) *hold with* e_{ni} *having the* d.f. F_{ni} *given by* (26), $1 \le i \le n$. *Suppose that* $\mathbf{X}$ *satisfies* (NX); (F_0, G) *and* (F_1, G) *satisfy* (5.5.68) – (5.5.70) *and that*

$$\int |F_1 - F_0|\, dG < \infty, \tag{27}$$

(a) *If, in addition,* (5.6a.10) *and* (5.6a.11) *hold, then*

$$K_1^* = \int \Big\{W_1^o + m_1 - f_0 \frac{\int (W_1^o + m_1)f_0\, dG}{\int f_0^2\, dG}\Big\}^2 dG + o_p(1) \tag{28}$$

provided

$$n^{-1/2} \Sigma_i\, \delta_{ni} = O(1). \tag{29}$$

(b) *Without any additional conditions,*

(30) $$K_2^* = \int \|\mathbf{W}^o + \mathbf{m}_2 - f_0 \frac{\int(\mathbf{W}^o + \mathbf{m}_2) f_0 dG}{\int f_0^2 dG}\|^2 dG + o_p(1),$$

provided

(31) $$\Sigma_i \mathbf{A}\mathbf{x}_{ni}\delta_{ni} = O(1).$$

Proof. Apply Theorem 5.5.1 and (5.5.49) to $D = n^{-1/2}[1, 0, \ldots, 0]$, $Y_{ni} \equiv e_{ni}$, $H_{ni} \equiv F_0$, $\{F_{ni}\}$ given by (26) to conclude (a). Apply the same results to $D = \mathbf{AX}$ and the rest of the entities as in the proof of (a) to conclude (b). □

Now apply Lemma 6.3.1 to

(30) $$Z_n := W^o + m_1 - f_0 \frac{\int(W^o + m_1)f_0 dG}{\int f_0^2 dG},$$
$$Z := B(F_0) + a_1(F_1 - F_0) - f_0 \frac{\int\{B(F_0) + a_1(F_1 - F_0)\}f_0 dG}{\int f_0^2 dG},$$

where $a_1 := \limsup_n n^{-1/2} \Sigma_i \delta_{ni}$, to obtain

Corollary 6.3.4. *Under the conditions of Lemma* 6.3.2(a),

$$K_1^* \xrightarrow[d]{} \int Z^2 dG, \text{ where } Z \text{ is as in } (30). \qquad □$$

Similarly, apply Lemma 6.3.1 to

(31) $$\mathbf{Z}_n := \mathbf{W}^o + \mathbf{m}_2 - f_0 \frac{\int(\mathbf{W}^o + \mathbf{m}_2)f_0 dG}{\int f_0^2 dG},$$
$$\mathbf{Z} := B(F_0) + \mathbf{a}_2(F_1 - F_0) - f_0 \frac{\int\{B(F_0) + \mathbf{a}_2(F_1 - F_0)\}f_0 dG}{\int f_0^2 dG},$$

where $\mathbf{a}_2 = \limsup_n \Sigma_i \mathbf{A}\mathbf{x}_{ni}\delta_{ni}$, to obtain

Corollary 6.3.5. *Under the conditions of Lemma* 6.3.2(b),

$$K_2^* \xrightarrow[d]{} \int \|\mathbf{Z}\|^2 dG, \text{ where } \mathbf{Z} \text{ is as in } (31). \qquad □$$

An interesting choice of $\delta_{ni} = p^{-1/2}\|\mathbf{A}\mathbf{x}_{ni}\|$. Another choice is $\delta_{ni} \equiv n^{-1/2}$. Both a priori satisfy (26), (29) and (31). □

6.4. TESTING WITH UNKNOWN SCALE

Now consider (1.1.1) and the problem of testing H_1 of (6.1.4). Here we shall discuss the modifications of $\hat{D}_j$, $\hat{K}_j$, $j = 1, 2$, of Sections 6.2, 6.3 that will be suitable for H_1. With W_1^o, $\mathbf{W}^o$ as before, define

$$\begin{aligned}
(1)\qquad & D_1(a, \mathbf{u}) := \sup_y |W_1^o(ay, \mathbf{u})|, \\
& D_2(a, \mathbf{u}) := \sup_y |\mathbf{W}^o(ay, \mathbf{u})|, \\
& K_1(a, \mathbf{u}) := \int \{W_1^o(ay, \mathbf{u})\}^2 dG(y), \\
& K_2(a, \mathbf{u}) := \int \|\mathbf{W}^o(ay, \mathbf{u})\|^2 dG(y), \qquad a > 0,\ \mathbf{u} \in \mathbb{R}^p.
\end{aligned}$$

Let $(\tilde{\sigma}, \tilde{\beta})$ be estimators of (σ, β), $\tilde{D}_j$ and $\tilde{K}_j$ stand for $D_j(\tilde{\sigma}, \tilde{\beta})$ and $K_j(\tilde{\sigma}, \tilde{\beta})$, respectively, $j = 1, 2$. The following two theorems give the a.n.d.'s of these statistics. Theorem 6.4.1 follows from Corollary 2.3.4 in a similar fashion as does Theorem 6.2.1 from Corollaries 2.3.3 and 2.3.5. Theorem 6.4.2 follows from Theorems 5.5.8 in a similar fashion as does Theorem 6.3.2 from Theorem 5.5.1. Recall the conditions $(\mathbf{F}_0\mathbf{1})$ and $(\mathbf{F}_0\mathbf{3})$ from Section 2.3.

Theorem 6.4.1. *In addition to* (1.1.1) *and* H_1, *assume that* (NX), $(\mathbf{F}_0\mathbf{1})$, $(\mathbf{F}_0\mathbf{3})$ *and the following hold.*

$$(2)\quad (a)\ |n^{1/2}(\tilde{\sigma} - \sigma)\sigma^{-1}| = O_p(1). \quad (b)\ \|\mathbf{A}^{-1}(\tilde{\beta} - \beta)\| = O_p(1).$$

Then,

$$\tilde{D}_1 = \sup |W_1(t) + q_0(t)\{n^{1/2}\bar{\mathbf{x}}_n'(\tilde{\beta}-\beta) + n^{1/2}(\tilde{\sigma}-\sigma)F_0^{-1}(t)\}\sigma^{-1}| + o_p(1),$$

and

$$\tilde{D}_2 = \sup \|\mathbf{W}(t) + q_0(t)\{\mathbf{A}^{-1}(\tilde{\beta}-\beta) + n^{1/2}\mathbf{A}\bar{\mathbf{x}}_n \cdot n^{1/2}(\tilde{\sigma}-\sigma)F_0^{-1}(t)\}\sigma^{-1}\| + o_p(1),$$

where now $W_1(\cdot) := W_1^o(\sigma F_0^{-1}(\cdot), \beta)$ *and* $\mathbf{W}(\cdot) := \mathbf{W}^o(\sigma F_0^{-1}(\cdot), \beta)$.

Theorem 6.4.2. *In addition to* (1.1.1) *and* H_1, *assume that* (NX), (2), (5.5.69) *with* $F = F_0$, *and the following hold.*

(3) F_0 *has a continuous density* f_0 *such that*

(a) $0 < \int |y|^j f_0^k(y) dG(y) < \infty$, $j = 0, k = 1, 2$; $j = 2, k = 2$.

(b) $\lim_{s\to 0} \limsup_n \int f_0^k(y+\tau n^{-1/2}+s)dG(y) = \int f_0^k\, dG(y)$, $k = 1, 2$, $\tau \in \mathbb{R}$.

(c) $\lim_{s\to 0} \int |y|\, f_0(y(1+s))dG(y) = \int |y|\, f_0(y)dG(y)$.

Then,

$$\tilde{K}_1 = \int [W_1^o(\sigma y, \beta) + f_0(y)\{n^{1/2}\bar{x}_n'(\tilde{\beta} - \beta) + n^{1/2}(\tilde{\sigma} - \sigma)y\}\sigma^{-1}]^2 dG(y) + o_p(1),$$

$$\tilde{K}_2 = \int \|W^o(\sigma y, \beta) + f_0(y)\{A^{-1}(\tilde{\beta} - \beta) + n^{1/2}A\bar{x}_n \cdot n^{1/2}(\tilde{\sigma} - \sigma)y\}\sigma^{-1}\|^2 dG(y) + o_p(1).$$

Clearly, from these theorems one can obtain an analogue of Corollary 6.3.2 when $(\tilde{\sigma}, \tilde{\beta})$ are chosen to be asymptotically efficient estimators.

As is the case in the classical least square theory or in the M-estimation methodology, neither of the two dispersions $K_1(a, \mathbf{u})$ and $K_2(a, \mathbf{u})$ can be used to satisfactorily estimate (σ, β) by the simultaneous minimization process. The analogues of the m.d. goodness-of-fit tests that should be used are $\inf\{K_j(\tilde{\sigma}, \mathbf{u});\ \mathbf{u} \in \mathbb{R}^p\}$, $j = 1, 2$. The methodology of Section 5 may be used to obtain the asymptotic distributions of these statistics in a fashion similar to the above. □

6.5. TESTING FOR SYMMETRY OF THE ERRORS

Consider the model (1.1.1) and the hypothesis H_s of symmetry of the errors specified at (6.1.5). The proposed tests are to be based on $\hat{D}_{js}$, $j = 1, 2, 3$, of (6.1.6), (6.1.7), $K_j^+(\hat{\beta})$, and $\inf\{K_j^+(\mathbf{t});\ \mathbf{t} \in \mathbb{R}^p\}$, $j = 1, 2$, where

$$(1) \qquad K_1^+(\mathbf{t}) := \int \{W_1^+(y, \mathbf{t})\}^2 dG(y), \quad K_2^+(\mathbf{t}) := \int \|\mathbf{W}^+(y, \mathbf{t})\|^2 dG(y), \quad \mathbf{t} \in \mathbb{R}^p,$$

with W_1^+ and $\mathbf{W}^+$ as in (6.1.7) and (6.1.9). Large values of these statistics are considered to be significant for H_s.

Although the results of Chapters 2 and 5 can be used to obtain their asymptotic behavior under fairly general alternatives, here we shall focus only on the a.n.d.'s of these tests. To state these, we need some more notation. For a d.f. F, define

$$(2) \qquad F_+(y) := F(y) - F(-y), \qquad y \geq 0.$$

Then, with F^{-1} denoting the usual inverse of a d.f. F, we have

$$(3)\qquad F_+^{-1}(t) = F^{-1}((1+t)/2), \quad -F_+^{-1}(t) = F^{-1}((1-t)/2), \quad 0 \le t \le 1,$$

for all F that are continuous and symmetric around 0. Finally, let

$$(4)\qquad W_1^*(t) := W_1^+(F_+^{-1}(t), \beta), \quad \mathbf{W}^*(t) := \mathbf{W}^+(F_+^{-1}(t), \beta),$$
$$q^+(t) := f(F_+^{-1}(t)), \qquad 0 \le t \le 1.$$

We are now ready to state and prove

Theorem 6.5.1. *In addition to* (1.1.1), H_s *and* (**NX**), *assume that* F *in* H_s *and the estimator* $\hat{\beta}$ *satisfy* (**F1**) *and*

$$(5)\qquad \|\mathbf{A}^{-1}(\hat{\beta} - \beta)\| = O_p(1), \qquad \textit{under } H_s.$$

Then,

$$(6)\qquad \hat{D}_{1s} = \sup_{0\le t\le 1} |W_1^*(t) + 2q^+(t)\, n^{1/2}\bar{\mathbf{x}}_n'\mathbf{A}\, \mathbf{A}^{-1}(\hat{\beta} - \beta)| + o_p(1),$$

$$(7)\qquad \hat{D}_{2s} = \sup_{0\le t\le 1} |\mathbf{W}^*(t) + 2q^+(t)\, \mathbf{A}^{-1}(\hat{\beta} - \beta)| + o_p(1).$$

and

$$(8)\qquad \hat{D}_{3s} = \sup_{0\le t\le 1} \|\mathbf{W}^*(t) + 2q^+(t)\, \mathbf{A}^{-1}(\hat{\beta} - \beta)\| + o_p(1).$$

Proof. The proof follows from Theorem 2.3.1 in the following fashion. The details will be given only for (8), as they are the same for (7) and quite similar for (6). Because F is continuous and symmetric around 0 and because $\mathbf{W}^+(\cdot, \cdot) \equiv \mathbf{W}^+(-\cdot, \cdot)$, $\hat{D}_{3s} =_d \sup_{0\le t\le 1} \mathbf{W}^+(F_+^{-1}(t), \hat{\beta})$. But, from the definition (6.1.8) and (3), it follows that for a $\mathbf{v} \in \mathbb{R}^p$,

$$\mathbf{W}^+(F_+^{-1}(t), \mathbf{v})$$
$$= \Sigma_i\, \mathbf{A}\mathbf{x}_{ni}\{I(e_{ni} \le F^{-1}(\tfrac{1+t}{2}) + \mathbf{c}_{ni}'\mathbf{u}) + I(e_{ni} \le F^{-1}(\tfrac{1-t}{2}) + \mathbf{c}_{ni}'\mathbf{u}) - 1\}$$

$$(9)\qquad = \mathbf{S}(\tfrac{1+t}{2}, \mathbf{u}) + \mathbf{S}(\tfrac{1-t}{2}, \mathbf{u}) - \Sigma_i\, \mathbf{A}\mathbf{x}_{ni}, \qquad 0 \le t \le 1,$$

where

$$\mathbf{S}(t, \mathbf{u}) := \Sigma_i\, \mathbf{A}\mathbf{x}_{ni}\, I(e_{ni} \le F^{-1}(t) + \mathbf{c}_{ni}'\mathbf{u}), \qquad 0 \le t \le 1,$$

is a p–vector of S_d–processes of (2.3.1) with $X_{ni} \equiv e_{ni}$, $F_{ni} \equiv F \equiv H$, $\mathbf{c}_{ni} \equiv$

$\mathbf{A}\mathbf{x}_{ni}$, $\mathbf{u} = \mathbf{A}^{-1}(\mathbf{v} - \boldsymbol{\beta})$ and where the j^{th} process has the weights $\{d_{ni}\}$ given by the j^{th} column of $\mathbf{AX}$. The assumptions about F and $\mathbf{X}$ imply all the assumptions of Theorem 2.3.1. Hence (8) follows from (3.2.6), (5) and (9) in an obvious fashion. □

Next, we state an analogous result for the L_2-distances.

Theorem 6.5.2. *In addition to* (1.1.1), H_s, (NX) *and* (5), *assume that* F *in* H_s *and the integrating measure* G *satisfy* (5.3.8), (5.5.68), (5.5.70) *and* (5.6a.13). *Then,*

$$(10) \quad K_1^+(\hat{\boldsymbol{\beta}}) = \int [W_1^+(y) + 2f(y)\, n^{1/2}\bar{\mathbf{x}}_n'(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})]^2 dG(y) + o_p(1),$$

$$(11) \quad K_2^+(\hat{\boldsymbol{\beta}}) = \int \|\mathbf{W}^+(y) + 2f(y)\, \mathbf{A}^{-1}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})\|^2 dG(y) + o_p(1),$$

where $W_1^+(\cdot)$, $\mathbf{W}^+(\cdot)$ *now stand for* $W_1^+(\cdot, \boldsymbol{\beta})$, $\mathbf{W}^+(\cdot, \boldsymbol{\beta})$.

Proof. The proof follows from two applications of Theorem 5.5.2, once with $\mathbf{D} = n^{-1/2}[1, 0, ..., 0]$ and once with $\mathbf{D} = \mathbf{XA}$. In both cases, take Y_{ni} and F_{ni} of that theorem to be equal to e_{ni} and F, $1 \le i \le n$, respectively. The Claim 5.5.2 justifies the applicability of that theorem under the present assumptions. □

The next result is useful in obtaining the a.n.d.'s of the m.d. test statistics. Its proof uses Theorem 5.5.2 and 5.5.4 in a similar fashion as Theorems 5.5.1 and 5.5.3 are used in the proof of Theorem 6.3.1, and hence no details are given. Let

$$K_j^s := \inf\{K_j^+(\mathbf{t});\ \mathbf{t} \in \mathbb{R}^p\}, \quad j = 1, 2.$$

Theorem 6.5.3. *Assume that* (1.1.1), H_s, (NX), (5.3.8), (5.5.68), (5.5.70) *and* (5.6a.13) *hold.*

(a) *If, in addition,* (5.6a.10) *and* (5.6a.11) *hold, then*

$$(12) \quad K_1^s = 2\int_0^\infty \{W_1^+(y) - f(y)\int_0^\infty W_1^+ f dG\ (\int_0^\infty f^2 dG)^{-1}\}^2 dG + o_p(1).$$

(b) *Under no additional assumptions,*

$$(13) \quad K_2^s = 2\int_0^\infty \|\mathbf{W}^+(y) - f(y)\int_0^\infty \mathbf{W}^+ f dG\ (\int_0^\infty f^2 dG)^{-1}\|^2 dG + o_p(1). \qquad □$$

To obtain the a.n.d.'s of the given statistics from the above theorem we now apply Lemma 6.3.1 to the approximating processes. The details will be given for K_2^s only as they are similar for K_1^s. Accordingly, let

$$Z_n(y) := W^+(y) - f(y)\int_0^\infty W^+ f dG \left(\int_0^\infty f^2 dG\right)^{-1}, \quad n \ge 1, \; y \ge 0. \tag{14}$$

To determine the approximating r.v. for K_2^s we shall first obtain the covariance matrix function for this Z_n, the computation of which is made easy by rewriting Z_n as follows.

Recall the definition of ψ from (5.6a.2) and define

$$\alpha_i(y) := I(e_i \le y) + I(e_i \le -y) - 1, \;\; y \in \mathbb{R}, \qquad \bar{\alpha}_i := \int_0^\infty \alpha_i \, d\psi, \;\; 1 \le i \le n;$$

$$\boldsymbol{\alpha}' := (\alpha_1, \ldots, \alpha_n); \qquad \bar{\boldsymbol{\alpha}}' := (\bar{\alpha}_1, \ldots., \bar{\alpha}_n); \qquad a := \int_0^\infty f^2 dG.$$

Then

$$Z_n(y) = \mathbf{A}\mathbf{X}'[\boldsymbol{\alpha}(y) - f(y)\bar{\boldsymbol{\alpha}}\, a^{-1}], \qquad y \ge 0. \tag{15}$$

Now observe that under H_s, $E\boldsymbol{\alpha} = 0$, $E\alpha_1(x)\alpha_1(y) = 2\,(1-F(y))$, $0 \le x \le y$, and, because of the independence of the errors,

$$E\boldsymbol{\alpha}(x)\boldsymbol{\alpha}'(y) = 2(1-F(y))\, I_{p\times p}, \qquad 0 \le x \le y. \tag{16}$$

Again, because of the symmetry and the continuity of F and Fubini, for $y \ge 0$,

$$E\alpha_1(y)\bar{\alpha}_1 = \int_0^\infty E[I(e_1 \le y) + I(e_1 \le -y) - 1][I(e_1 \le x) + I(e_1 \le -x) - 1]\, d\psi(x)$$

$$= \int_0^\infty [F(x\wedge y) + F(-x\wedge y) - F(y) + F(x\wedge -y) + F(-x\wedge -y) - F(-y)]\, d\psi(x)$$

$$= 2(1-F(y))\{\psi(y) - \psi(0)\} + \int_y^\infty 2(1-F(x))\, d\psi(x)$$

$$= 2\int_y^\infty [\psi(x) - \psi(0)]\, dF(x) =: k(y), \quad \text{say.}$$

The last equality is obtained by integrating the second expression in the previous one by parts. From this and the independence of the errors, we obtain

$$E\boldsymbol{\alpha}(y)\bar{\boldsymbol{\alpha}}' = k(y)\, I_{p\times p}, \qquad y \ge 0.$$

Similarly,

$$E\bar{\alpha}\bar{\alpha}' = I_{pxp}\, 4\int_0^\infty \int_x^\infty (1-F(y))\, d\psi(x)d\psi(y) =: I_{pxp}\, r(F,G), \text{ say.}$$

From these calculations one readily obtains that under H_s, for $0 \le x \le y$,

$$\text{(17)} \qquad K_n(x, y) := EZ_n(x)Z_n'(y)$$
$$= [2(1-F(y)) - k(y)f(x)a^{-1} - k(x)f(y)a^{-1} + r(F,G)]I_{pxp}.$$

We also need the weak convergence of W^+ to a continuous Gaussian process in uniform topology. One way to prove this is as follows. By (16),

$$\text{(18)} \qquad EW^+(x)W^+(y)' = 2(1 - F(y))\, I_{pxp}, \qquad 0 \le x \le y,$$

From the definition (6.1.9) and the symmetry of F,

$$\text{(19)} \qquad W^+(y) = \Sigma_i\, Ax_{ni}\{I(e_{ni} \le y) - I(-e_{ni} < y)\}$$
$$= \Sigma_i\, Ax_{ni}\{I(e_{ni} \le y) - F(y)\} - \Sigma_i\, Ax_{ni}\{I(-e_{ni} \le y) - F(y)\}$$
$$+ \Sigma_i\, Ax_{ni}I(-e_{ni} = y)$$
$$\text{(20)} \qquad = \mathcal{W}_1(y) - \mathcal{W}_2(y) + \Sigma_i\, Ax_{ni}I(-e_{ni} = y)\,, \text{ say,} \qquad y \ge 0.$$

Now, let $\mathbb{W}' := (\mathbb{W}_1, \mathbb{W}_2, ..., \mathbb{W}_p)$ be a vector of independent Wiener processes on [0, 1] such that $\mathbb{W}(0) = 0$, $E\mathbb{W} \equiv 0$, and $E\mathbb{W}_j(s)\mathbb{W}_j(t) = s\wedge t$, $1 \le j \le p$. Note that

$$E\mathbb{W}(2(1-F(x)))\mathbb{W}(2(1-F(y)))' = 2(1-F(y))\, I_{pxp}, \qquad 0 \le x \le y.$$

From (18) and (19), it hence follows, with the aid of the L–F CLT and the Cramer–Wold device, that under (NX), all finite dimensional distributions of W^+ converge to those of $\mathbb{W}(2(1-F))$.

To prove the tightness in the uniform metric, proceed as follows. From (20) and the triangle inequality, because of (NX), it suffices to show that $\mathcal{W}_1$ and $\mathcal{W}_2$ are tight. But by the symmetry and the continuity of F,

$$\{\mathcal{W}_1(y), y \in \mathbb{R}\} =_d \{\mathcal{W}_2(y), y \in \mathbb{R}\} =_d \{\mathcal{W}_1(F^{-1}(t)), 0 \le t \le 1\}.$$

But, $\mathcal{W}_1(F^{-1})$ is obviously a p-vector of w.e.p.'s of the type W_d^* specified at (2.2a.33). Thus the tightness follows from (2.2a.35) of Corollary 2.2a.1. We summarize this weak convergence result as

Lemma 6.5.1. *Let* F *be a continuous d.f. that is symmetric around* 0 *and* $\{e_{ni}, 1 \le i \le n\}$ *be i.i.d.* F *r.v.'s. Assume that* (NX) *holds. Then,*

$$W^+(\cdot) \Rightarrow \mathscr{W}(2(1-F(\cdot))) \text{ in } (\mathbb{D}[0,\infty], d). \qquad \square$$

The above discussion suggests the approximating process for the $\mathbf{Z}_n$ of (16) to be

$$\mathbf{Z}(y) := \mathscr{W}(2(1-F(y)))-f(y)\int_0^\infty \mathscr{W}(2(1-F))f dG \left(\int_0^\infty f^2 dG\right)^{-1}, \quad y \geq 0. \tag{21}$$

Straightforward calculations show that $\mathbf{K}_n(x, y) = E\mathbf{Z}(x)\mathbf{Z}'(y)$, $0 \leq x \leq y$, $n \geq 1$. This then verifies (i), (ii) and (iv) of Lemma 6.3.1 in the present case. Condition (iii) is verified as in the proof of Corollary 6.3.1(b) with the help of Lemma 6.5.1. To summarize, we have

Corollary 6.5.1. (a) *Under the conditions of Theorem* 6.5.3(a),

$$K_1^s \xrightarrow[d]{} 2\int_0^\infty [\mathscr{W}_1(2(1-F(y)))-f(y)\int_0^\infty \mathscr{W}_1(2(1-F))f dG \left(\int_0^\infty f^2 dG\right)^{-1}]^2 dG(y). \tag{22}$$

(b) *Under the conditions of Theorem* 6.5.3(b),

$$K_2^s \xrightarrow[d]{} 2\int_0^\infty \|\mathbf{Z}\|^2 dG(y), \quad \textit{with } \mathbf{Z} \textit{ given at } (21). \qquad \square \tag{23}$$

Remark 6.5.1. The distributions of the limiting r.v.'s in (22) and (23) have been studied by Martynov (1975, 1976) and Boos (1982) for some F and G. An interesting G in the present case is $G = \lambda$. But the corresponding tests are not a.d.f.. Also because the F in H_s is unknown, one can not use $G = F$ or the Anderson–Darling integrating measures $dG = dF/\{F(1-F)\}$ in these test statistics.

One way to overcome this problem would be to use the signed rank analogues of the above tests which is equivalent to replacing the F in the integrating measure by an appropriate empirical of the residuals $\{Y_{nj}-\mathbf{x}_{nj}'\mathbf{u};\ 1 \leq j \leq n\}$. Let R_{iu}^+ denote the rank of $|Y_{ni}-\mathbf{x}_{ni}'\mathbf{u}|$ among $\{|Y_{nj}-\mathbf{x}_{nj}'\mathbf{u}|;\ 1 \leq j \leq n\}$, $1 \leq i \leq n$, and define

$$\mathscr{Z}_1^+(t, \mathbf{u}) := n^{-1/2}\, \Sigma_i\, I(R_{iu}^+ \leq nt)\, \mathrm{sgn}(Y_{ni}-\mathbf{x}_{ni}'\mathbf{u}),$$

$$\mathscr{Z}_2^+(t, \mathbf{u}) := \mathbf{A}\, \Sigma_i\, \mathbf{x}_{ni}\, I(R_{iu}^+ \leq nt)\, \mathrm{sgn}(Y_{ni}-\mathbf{x}_{ni}'\mathbf{u}), \qquad 0 \leq t \leq 1,\ \mathbf{u} \in \mathbb{R}^p.$$

The signed rank analogues of K_1^s, K_2^s statistics, respectively, are $\mathcal{K}_1^s := \inf\{\mathcal{K}_1(\mathbf{u});\ \mathbf{u} \in \mathbb{R}^p\}$, $\mathcal{K}_2^s := \inf\{\mathcal{K}_2(\mathbf{u});\ \mathbf{u} \in \mathbb{R}^p\}$, where

$$\mathcal{K}_1(\mathbf{u}) := \int_0^1 [\mathscr{Z}_1^+(t, \mathbf{u})]^2 dL(t), \qquad \mathcal{K}_2(\mathbf{u}) := \int_0^1 \|\mathscr{Z}_2^+(t, \mathbf{u})\|^2 dL(t), \qquad \mathbf{u} \in \mathbb{R}^p,$$

with $L \in \mathbb{D}\mathcal{I}[0, 1]$. If $L(t) \equiv t$ then $\mathcal{K}_j^s$, $j = 1, 2$, are analogues of the Cramer–Von Mises statistics. If L is specified by the relation $dL(t) = \{1/t(1-t)\}dt$, then the corresponding tests would be the Anderson–Darling type test of symmetry.

Note that if in (3.3.1) we put $d_{ni} \equiv n^{-1/2}$, $X_{ni} \equiv e_{ni}$, $F_{ni} \equiv F$, then Z_d^+ of (3.3.1) reduces to $\mathcal{Z}_1^+$. Similarly, $\mathcal{Z}_2^+$ corresponds to a p-vector of Z_d^+-processes of (3.3.1) whose j^{th} component has $d_{ni} \equiv (j^{th}$ column of $\mathbf{A})'\mathbf{x}_{ni}$ and the rest of the entities the same as above. Consequently, from (3.3.17) and arguments like those used for Theorem 6.5.3, we can deduce the following

Theorem 6.5.4. *Assume that* (1.1.1), H_s *and* (**NX**) *hold;* L *is a d.f. on* $[0, 1]$, *and* F *of* H_s *satisfies* (**F1**), (**F2**).

(a) *If, in addition,* (5.6a.10) *and* (5.6a.11) *hold, then*

$$\mathcal{K}_1^s \xrightarrow[d]{} \int_0^1 [\mathcal{W}_1(t) - q^+(t)\int_0^1 \mathcal{W}_1\, q^+ dL\, (\int_0^1 (q^+)^2 dL)^{-1}]^2 dL(t). \tag{24}$$

(b) *Under no additional assumptions,*

$$\mathcal{K}_2^s \xrightarrow[d]{} \int_0^1 \|\mathcal{W}(t) - q^*(t)\int_0^1 \mathcal{W}\, q^* dL\, (\int_0^1 (q^*)^2 dL)^{-1}\|^2 dL(t), \tag{25}$$

where $q^*(t) := 2[f(F^{-1}((t+1)/2) - f(0)]$, $0 \leq t \leq 1$. □

Clearly this theorem covers $L(t) \equiv t$ case but not the case where $dL(t) = \{1/t(1-t)\}dt$. The *problem of proving an analogue of the above theorem for a general* L is *unsolved* at the time of this writing. □□

CHAPTER 7

AUTOREGRESSION

7.1. INTRODUCTION.

The purpose of this chapter is to offer a unified functional approach to some aspects of robust estimation and goodness-of-fit testing problems in pth order autoregression (AR(p)) models. This approach is similar to that of the previous chapters in connection with linear regression models, thereby extending a statistical methodology to one of the most applied models with dependent observations.

As before, let F be a d.f. on $\mathbb{R}$, $p \geq 1$ be an integer, $\epsilon_1, \epsilon_2, \ldots$ be i.i.d. F r.v.'s and $\mathbf{Y}_0 := (X_0, X_{-1}, \ldots, X_{1-p})'$ be an observable random vector independent of $\epsilon_1, \epsilon_2, \ldots$. In an AR(p) model one observes $\{X_i\}$ satisfying

$$(1) \qquad X_i = \rho_1 X_{i-1} + \cdots + \rho_p X_{i-p} + \epsilon_i, \qquad 1 \leq i \leq n, \ p \geq 1,$$

for some $\boldsymbol{\rho}' = (\rho_1, \rho_2, \ldots, \rho_p) \in \mathbb{R}^p$.

Processes that play a fundamental role in the robust estimation of $\boldsymbol{\rho}$ in this model are the *randomly weighted residual empirical* processes

$$(2) \qquad T_j(x, \mathbf{t}) := n^{-1} \sum_{i=1}^{n} g(X_{i-j})\, I(X_i - \mathbf{t}'\mathbf{Y}_{i-1} \leq x), \ \ x \in \mathbb{R}, \mathbf{t} \in \mathbb{R}^p, 1 \leq j \leq p,$$

where g is a measurable function from $\mathbb{R}$ to $\mathbb{R}$ and $\mathbf{Y}_{i-1} := (X_{i-1}, \ldots, X_{i-p})'$, $1 \leq i \leq n$. Let $\mathbf{T} := (T_1, \ldots, T_p)'$.

The generalized M- (GM) estimators of $\boldsymbol{\rho}$, as proposed by Denby and Martin (1979), are solution $\mathbf{t}$ of the p equations

$$(3) \qquad \mathcal{G}_j(\mathbf{t}) := \int \psi(x)\, T_j(dx, \mathbf{t}) = 0, \qquad 1 \leq j \leq p,$$

where ψ is a nondecreasing bounded measurable function from $\mathbb{R}$ to $\mathbb{R}$. These estimators are analogues of M-estimators of $\boldsymbol{\beta}$ in linear regression as discussed in Capter 4. Note that taking $g(x) \equiv xI[|x| \leq k] + kI[|x| > k] \equiv \psi(x)$ in (3) gives the Huber(k) estimators and taking $g(x) \equiv x \equiv \psi(x)$ gives the famous least square estimator.

The m.d. estimator $\boldsymbol{\rho}_g^+$ that is an analogue of $\boldsymbol{\beta}_D^+$ of (5.2.20) is defined as a minimizer, w.r.t. $\mathbf{t}$, of

$$(4) \qquad K_g(\mathbf{t}) = \sum_{j=1}^{p} \int [n^{-1/2} \sum_{i=1}^{n} g(X_{i-j})\{I(X_i \leq x + \mathbf{t}'\mathbf{Y}_{i-1}) - I(-X_i < x - \mathbf{t}'\mathbf{Y}_{i-1})\}]^2\, dG(x), \quad \mathbf{t} \in \mathbb{R}^p.$$

Observe that K involves **T**. In fact, $\forall\ \mathbf{t} \in \mathbb{R}^p$,

$$K_g(\mathbf{t}) = \sum_{j=1}^{p} \int [n^{1/2}\{T_j(x, \mathbf{t}) - \sum_{i=1}^{n} g(X_{i-j}) + T_j(-x, \mathbf{t})\}]^2 \, dG(x).$$

Three members of this class of estimators are of special interest. They correspond to the cases $g(x) \equiv x \equiv G(x)$; $g(x) \equiv x$, $G \equiv \delta_0$, the measure degenerate at 0; $g(x) \equiv x$, $G \equiv F$ in the F known case. The first gives an analogue of the *Hodges–Lehmann* (h.l.) estimator of $\boldsymbol{\rho}$, the second gives the *least absolute deviation* (l.a.d.) estimator, while the third gives an estimator that is more efficient at logistic (double exponential) errors than l.a.d (h.l.) estimator.

Another important process in the model (1) is the ordinary residual empirical process

$$\text{(5)} \qquad F_n(x, \mathbf{t}) := n^{-1} \Sigma_i I(X_i - \mathbf{t}' \mathbf{Y}_{i-1} \le x), \qquad x \in \mathbb{R}, \ \mathbf{t} \in \mathbb{R}^p.$$

An estimator of F or a test of goodness-of-fit pertaining to F are usually based on $F_n(x, \hat{\boldsymbol{\rho}})$, where $\hat{\boldsymbol{\rho}}$ is an estimator of $\boldsymbol{\rho}$.

Clearly F_n is a special case of (2). But, both F_n and T_j, $1 \le j \le p$, are special cases of

$$\text{(6)} \quad W_h(x, \mathbf{t}) := n^{-1} \Sigma_i h(\mathbf{Y}_{i-1}) I(X_i - \mathbf{t}' \mathbf{Y}_{i-1} \le x)$$
$$= n^{-1} \Sigma_i h(\mathbf{Y}_{i-1}) I(\epsilon_i \le x + (\mathbf{t} - \boldsymbol{\rho})' \mathbf{Y}_{i-1}), \qquad x \in \mathbb{R}, \ \mathbf{t} \in \mathbb{R}^p,$$

where h is a measurable function from $\mathbb{R}^p$ to $\mathbb{R}$. Choosing $h(\mathbf{Y}_{i-1}) \equiv g(X_{i-j})$ in W_h gives T_j, $1 \le j \le p$ and the choice of $h \equiv 1$ yields F_n.

From the above discussion it is apparent that the investigation of the large sample behavior of various inferential procedures pertaining to $\boldsymbol{\rho}$ and F, based on $\{T_j\}$ and $F_n(\cdot, \hat{\boldsymbol{\rho}})$, is facilitated by the weak convergence properties of $\{W_h(x, \boldsymbol{\rho}+n^{-1/2}\mathbf{u}),\ x \in \mathbb{R},\ \mathbf{u} \in \mathbb{R}^p\}$. This will be investigated in Section 7.2, with the aid of Theorem 2.2b.1. In particular, this section contains an a.u.l. result about $\{W_h(x, \boldsymbol{\rho}+n^{-1/2}\mathbf{u}),\ x \in \mathbb{R},\ \|\mathbf{u}\| \le B\}$ which in turn yields a.u.l. results about $\{\mathbf{T}(x, \boldsymbol{\rho}+n^{-1/2}\mathbf{u}),\ x \in \mathbb{R},\ \|\mathbf{u}\| \le B\}$ and $\{F_n(x, \boldsymbol{\rho}+n^{-1/2}\mathbf{u}),\ x \in \mathbb{R},\ \|\mathbf{u}\| \le B\}$. These results are useful in studying GM- and R-estimators of $\boldsymbol{\rho}$, akin to Chapters 3 and 4 when dealing with linear regression models. They are also useful in studying the large sample behaviour of some tests of goodness-of-fit pertaining to F. Analogous results about the ordinary empirical of the residuals in autoregressive *moving average* models are briefly discussed in Remark 7.2.4.

Generalized M-estimators and analogues of Jaeckel's (1972) R-estimators are discussed in Section 7.3. In order to use R- or m.d.

estimators to construct confidence intervals one often needs consistent estimators of the functional Q(f) of the error density f. Appropriate analogues of estimators of Q(f) of Section 4.5 are shown to be consistent under (**F1**) and (**F2**). This is also done in Section 7.3, with the help of the a.u.l. property of $\{F_n(x, \rho + n^{-1/2}\mathbf{u}),\ x \in \mathbb{R},\ \|\mathbf{u}\| \le B\}$. This result is also used to prove the a.u.l. of serial rank correlations of the residuals in an AR(p) model. Such results should be useful in developing anlogues of the method of moment estimators or Yule-Walker equations based on ranks in AR(p) models.

Section 7.4 investigates the behaviour of two classes of m.d. estimators of ρ, including the class of estimators $\{\rho_g^+\}$. A crucial result needed to obtain the asymptotic distributions of these estimators is the asymptotic uniform quadraticity of their defining dispersions. This result is also proved in Section 7.4. Section 7.5 contains appropriate analogues of some of the goodness-of-fit tests of Capter 6 pertaining to F.

7.2. ASYMPTOTIC UNIFORM LINEARITY OF $\mathcal{W}_h$ and F_n.

Recall the definition of V_h process from (1.4.1) and the statement of Theorem 2.2b.1. In (1.4.1), let

$$(1)\qquad \zeta_{ni} \equiv \epsilon_i,\ h_{ni} \equiv h(\mathbf{Y}_{i-1}),\ \delta_{ni} \equiv n^{-1/2}\mathbf{u}'\mathbf{Y}_{i-1}, \quad \mathbf{u} \in \mathbb{R}^p,\ 1 \le i \le n,$$

$$\mathcal{A}_{n1} = \sigma\text{--field } \{\mathbf{Y}_0\},\ \ \mathcal{A}_{ni} = \sigma\text{--field } \{\mathbf{Y}_0', \epsilon_1, \ldots, \epsilon_{i-1}\}, \quad 2 \le i \le n.$$

Then one readily sees that the corresponding $V_h(x)$, $V_h^*(x)$ are, respectively, equal to $\mathcal{W}_h(x, \rho + n^{-1/2}\mathbf{u})$, $\mathcal{W}_h(x, \rho)$ for each $\mathbf{u} \in \mathbb{R}^p$ and for all $x \in \mathbb{R}$. Consequently, if we let

$$(2)\qquad \nu_h(x, \mathbf{t}) := n^{-1} \Sigma_i\, h(\mathbf{Y}_{i-1})\, F(x + (\mathbf{t} - \rho)'\mathbf{Y}_{i-1}),$$

$$\mathscr{W}_h(x, \mathbf{t}) := n^{1/2}[\mathcal{W}_h(x, \mathbf{t}) - \nu_h(x, \mathbf{t})], \qquad x \in \mathbb{R},\ \mathbf{t} \in \mathbb{R}^p,$$

then the corresponding $U_h(x)$, $U_h^*(x)$ are, respectively, equal to $\mathscr{W}_h(x, \rho + n^{-1/2}\mathbf{u})$, $\mathscr{W}_h(x, \rho)$ for each $\mathbf{u} \in \mathbb{R}^p$ and for all $x \in \mathbb{R}$. Recall the conditions (**F1**) and (**F2**) from Corollary 2.3.1. We are now ready to state and prove the following

Theorem 7.2.1. *In addition to* (7.1.1), *assume that the following conditions hold:*

(a1) h *is a bounded function.*

(a2) $\qquad n^{-1/2} \max_{1 \le i \le n} \|Y_{i-1}\| = o_p(1).$

(a3) $\qquad n^{-1} \sum_{i=1}^{n} \|h(Y_{i-1})Y_{i-1}\| = O_p(1).$

(a4) $\qquad$ F *satisfies* **(F1)** *and* **(F2)**.

Then, for every $0 < B < \infty$,

$$\sup_{x \in \mathbb{R}, \|u\| \le B} |\mathscr{W}_h(x, \rho + n^{-1/2}u) - \mathscr{W}_h(x, \rho)| = o_p(1), \tag{3}$$

and

$$n^{1/2}[W_h(x, \rho + n^{-1/2}u) - W_h(x, \rho)] = -u' n^{-1} \Sigma_i h(Y_{i-1}) Y_{i-1} f(x) + \bar{o}_p(1). \tag{4}$$

where $\bar{o}_p(1)$ *is a sequence of stochastic processes that converges to zero, uniformly over the set* $x \in \mathbb{R}$, $\|u\| \le B$, *in probability.*

Proof. In view of the discussion preceeding the statement of the theorem it is clear that (2.2b.2) of Theorem 2.2b.1 applied to entities given in (1) above readily yields that

$$\sup\{|\mathscr{W}_h(x, \rho + n^{-1/2}u) - \mathscr{W}_h(x, \rho)|;\, x \in \mathbb{R}\} = o_p(1) \quad \text{for every fixed } u \in \mathbb{R}^p.$$

It is the uniformity with respect to u that requires an extra argument and that also turns out to be a consequence of another application of (2.2b.2) and a monotonic property inherent in these processes as we now show.

Since h is fixed, it will not be exhibited in the proof. Also, for convenience, write $\mathscr{W}(\cdot)$, $\mathscr{W}_u(\cdot)$, $W_u^{\pm}(\cdot)$, $\nu_u^{\pm}(\cdot)$ etc. for $\mathscr{W}_h(\cdot, \rho)$, $\mathscr{W}_h(\cdot, \rho + n^{-1/2}u)$, $W_h^{\pm}(\cdot, \rho + n^{1/2}u)$, $\nu_h^{\pm}(\cdot, \rho + n^{-1/2}u)$ etc. with $\pm$ signifying the fact that $h^{\pm}$ now appears in the place of h in these processes where $h^+ = 0 \vee h$, $h^- = h - h^+$. To avoid displays being broken into different lines often, write ξ_i, h_i, $h_i^{\pm}$ for Y_{i-1}, $h(Y_{i-1})$, $h^{\pm}(Y_{i-1})$, respectively, $i \ge 1$. Thus, e.g.,

$$\mathscr{W}_u^{\pm}(x) = n^{-1/2} \Sigma_i h_i^{\pm} \{I(\epsilon_i \le x + n^{-1/2}u'\xi_i) - F(x + n^{-1/2}u'\xi_i\}. \tag{5}$$

We also need the following processes:

$$\begin{aligned} T^{\pm}(x; u, a) &:= n^{-1/2} \Sigma_i h_i^{\pm} I(\epsilon_i \le x + n^{-1/2}u'\xi_i + n^{-1/2}a\|\xi_i\|), \\ m^{\pm}(x; u, a) &:= n^{-1/2} \Sigma_i h_i^{\pm} F(x + n^{-1/2}u'\xi_i + n^{-1/2}a\|\xi_i\|) \\ Z^{\pm} &:= T^{\pm} - m^{\pm}, \qquad\qquad x \in \mathbb{R},\ u \in \mathbb{R}^p,\ a \in \mathbb{R}. \end{aligned} \tag{6}$$

Observe that if in U_h of (2.2b.1) we take $\zeta_{ni} \equiv \epsilon_i$, $h_{ni} \equiv h^{\pm}(\xi_i)$, $\delta_{ni} \equiv n^{-1/2}\{u'\xi_i + a\|\xi_i\|\}$ and $\mathcal{A}_{ni}$, $1 \le i \le n$, as in (1), we obtain

$$U_h(\cdot) = Z^{\pm}(\cdot;\, u, a), \qquad \text{for every } u \in \mathbb{R}^p,\ a \in \mathbb{R}.$$

Similarly, if we take $\delta_{ni} \equiv n^{-1/2}u'\xi_i$ and the rest of the quantities as above then

$$U_h(\cdot) = Z^{\pm}(\cdot;\, u, 0) = \mathcal{W}_u^{\pm}(\cdot), \qquad \text{for every } u \in \mathbb{R}^p.$$

It thus follows from two applications of (2.2b.2) and the triangle inequality that for every $u \in \mathbb{R}^p$, $a \in \mathbb{R}$,

$$\text{(7a)} \qquad \sup_x |Z^{\pm}(x;\, u, a) - Z^{\pm}(x;\, u, 0)| = o_p(1),$$

$$\text{(7b)} \qquad \sup_x |\mathcal{W}_u^{\pm}(x) - \mathcal{W}^{\pm}(x)| = o_p(1).$$

Thus, to prove (3), because of the compactness of $\mathcal{N}(B)$, it suffices to show that for every $\epsilon > 0$ there is a $\delta > 0$ such that for every $\|u\| \le B$,

$$\text{(8)} \quad \limsup_n P\Big(\sup_{\|s\|\le B,\, \|s-u\|\le\delta,\, x} |\mathcal{W}_s(x) - \mathcal{W}_u(x)| > 4\epsilon\Big) < \epsilon.$$

By the definition of $\mathcal{W}^{\pm}$ and the triangle inequality, for $x \in \mathbb{R}$, $s, u \in \mathbb{R}^p$,

$$\text{(9)} \qquad |\mathcal{W}_s(x) - \mathcal{W}_u(x)| \le |\mathcal{W}_s^{+}(x) - \mathcal{W}_u^{+}(x, u)| + |\mathcal{W}_s^{-}(x) - \mathcal{W}_u^{-}(x)|,$$

$$|\mathcal{W}_s^{\pm}(x) - \mathcal{W}_u^{\pm}(x)| \le n^{1/2}[|W_s^{\pm}(x) - W_u^{\pm}(x)| + |\nu_s^{\pm}(x) - \nu_u^{\pm}(x)|].$$

But $s \in \mathcal{N}(B)$, $\|u\| \le B$, $\|s - u\| \le \delta$ imply that for all $1 \le i \le n$,

$$\text{(10)} \qquad n^{-1/2}u'\xi_i - n^{-1/2}\delta\|\xi_i\| \le n^{-1/2}s'\xi_i \le n^{-1/2}u'\xi_i + n^{-1/2}\delta\|\xi_i\|.$$

From (10), the monotonicity of the indicator function and the nonnegativity of $h^{\pm}$, we obtain

$$T^{\pm}(x;\, u, -\delta) - T^{\pm}(x;\, u, 0) \le W_s^{\pm}(x) - W_u^{\pm}(x) \le T^{\pm}(x;\, u, \delta) - T^{\pm}(x;\, u, 0)$$

for all $x \in \mathbb{R}$, $s \in \mathcal{N}(B)$, $\|s - u\| \le \delta$. Now center $T^{\pm}$ appropriately to obtain

(11) $$n^{1/2}|W_s^{\pm}(x)-W_u^{\pm}(x)|$$
$$\le |Z^{\pm}(x;\mathbf{u},\delta) - Z^{\pm}(x;\mathbf{u},0)| + |Z^{\pm}(x;\mathbf{u},-\delta) - Z^{\pm}(x;\mathbf{u},0)|$$
$$+ |m^{\pm}(x;\mathbf{u},\delta) - m^{\pm}(x;\mathbf{u},0)| + |m^{\pm}(x;\mathbf{u},-\delta) - m^{\pm}(x;\mathbf{u},0)|,$$

for all $x\in\mathbb{R}$, $\mathbf{s} \in \mathcal{N}(B)$, $\|\mathbf{s} - \mathbf{u}\| \le \delta$.

But, by (a4), $\forall\ \|\mathbf{u}\| \le B$,

(12) $$\sup_x |m^{\pm}(x;\mathbf{u},\pm\delta) - m^{\pm}(x;\mathbf{u},0)| \le \delta\|f\|_\infty n^{-1}\Sigma_i\|h_i\xi_i\|,$$

(13) $$\sup_{\|\mathbf{s}-\mathbf{u}\|\le\delta,x} n^{1/2}|\nu_s^{\pm}(x) - \nu_u^{\pm}(x)| \le \delta\|f\|_\infty n^{-1}\Sigma_i\|h_i\xi_i\|.$$

From (12), (11), (7a) applied with $a = \delta$ and $a = -\delta$ and the assumption (a3) one concludes that for every $\epsilon > 0$ there is a $\delta > 0$ such that for each $\|\mathbf{u}\| \le B$,

$$\limsup_n P\Big(\sup_{\|\mathbf{s}-\mathbf{u}\|\le\delta,x} n^{1/2}|W_s^{\pm}(x) - W_u^{\pm}(x)| > \epsilon\Big) \le \epsilon/2.$$

From this, (13), (9), and (a3) one now concludes (8) in a routine fashion. Finally, (4) follows from (3) and (a4) by Taylor's expansion of F. □

An application of (4) with $h(Y_{i-1}) = g(X_{i-j})$ and the rest of the quantities as in (1) readily yields the a.u.l. property of T_j–processes, $1\le j\le p$ of (7.1.2). This together with integration by parts yields the following expansion of the M-scores $\mathcal{G}_j$, $1\le j\le p$ of (7.1.3).

Corollary 7.2.1. *In addition to* (7.1.1), (a2) and (a4), *assume that the following conditions hold.*

(b1) g *is bounded.*

(b2) ψ *is nondecreasing, bounded and* $\int \psi\, dF = 0$.

(b3) $n^{-1}\Sigma_i \|g(X_{i-j})Y_{i-1}\| = O_p(1),\ 1 \le j \le p$.

Then, $\forall\ 0 < k, B < \infty$,

$$\sup|n^{1/2}[\mathcal{G}_j(\rho + n^{-1/2}\mathbf{u}) - \mathcal{G}_j(\rho)] - \mathbf{u}'n^{-1}\Sigma_i\, g(X_{i-j})Y_{i-1}\int f d\psi| = o_p(1)$$

where the supremum is taken over all ψ *with* $\|\psi\|_{tv} \le k < \infty$, $\|\mathbf{u}\| \le B$, $1\le j\le p$. □

Upon choosing $h \equiv 1$ in (4) one obtains an analogous result for the *ordinary residual empirical process* $F_n(x, \mathbf{t})$. Because of its importance and for an easy reference later on we state it as a separate result. Observe that in the following corollary the assumption (a3*) is nothing but the assumption (a3) of Theorem 7.2.1 with $h \equiv 1$.

Corollary 7.2.2. *Suppose that* (7.1.1) *holds. In addition, assume that* (a2), (a3*) *and* (a4) *hold, where,*

(a3*) $$n^{-1}\Sigma_i \|\mathbf{Y}_{i-1}\| = O_p(1).$$

Then, for every $0 < B < \infty$,

(14) $$\sup|n^{1/2}\{F_n(x, \rho+n^{-1/2}\mathbf{u}) - F_n(x, \rho)\} - \mathbf{u}' n^{-1}\Sigma_i \mathbf{Y}_{i-1} f(x)| = o_p(1),$$

where the supremum is taken over $x \in \mathbb{R}$, $\|\mathbf{u}\| \leq B$. □

Remark 7.2.1. Observe that non of the above results require that the process $\{X_i\}$ be stationary or any of the moments be finite. □

Remark 7.2.2. Consider the ***assumptions (a2) and (a3).*** If $\mathbf{Y}_0$ and $\{\epsilon_i\}$ are so chosen as to make $\{X_i\}$ stationary, ergodic and if $E(\|\mathbf{Y}_0\|^2+\epsilon_1^2)<\infty$ then (a2) is *a priori* satisfied and (a1) implies (a3). See, e.g., Anderson (1971; p 203). In paricular, (a3) holds for the h corresponding to the Huber function $h(x) \equiv |x|I(|x| \leq k) + \text{sign}(x)I(|x| > k)$, $k > 0$.

Of course if (a1) holds with the function h bounded in such a way that puts zero weight outside of compacts then (a3) is trivially satisfied.

Observe that (a2) is *weaker than requiring the finiteness of the second moment.* To see this, consider, for example, an AR(1) model where X_0 and $\epsilon_1, \epsilon_2, \ldots$ are independent r.v.'s and for some $|\rho| < 1$,

$$X_i = \rho X_{i-1} + \epsilon_i, \qquad i \geq 1.$$

Then,

$$X_i = \rho^i X_0 + \sum_{j=1}^{i} \rho^{j-i}\epsilon_j, \qquad \mathbf{Y}_i = X_i, \qquad i \geq 1.$$

Thus, here (a2) is implied by

(*i*) $$\max_{1\leq i\leq n} n^{-1/2}|\epsilon_i| = o_p(1).$$

But, (*i*) is equivalent to showing that $x^2 \ell n\,\{1-P(|\epsilon_1|>x)\} \longrightarrow 0$ as $x \longrightarrow \infty$, which, in turn is equivalent to requiring that $x^2P(|\epsilon_1|>x) \longrightarrow 0$ as $x \longrightarrow \infty$. This last condition is weaker than requiring that $E|\epsilon_1|^2 < \infty$. For example, let the right tail of the distribution of $|\epsilon_1|$ be given as follows:

$$P(|\epsilon_1|>x) = 1, \qquad x < 2,$$
$$= 1/(x^2 \ln x), \qquad x \geq 2.$$

Then, $E|\epsilon_1| < \infty$, $E\,\epsilon_1^2 = \infty$, yet $x^2 P(|\epsilon_1|>x) \longrightarrow 0$ as $x \longrightarrow \infty$. □

Remark 7.2.3. An analogue of (14) was first proved by Boldin (1982) requiring $\{X_i\}$ to be stationary, $E\epsilon_1 = 0$, $E(\epsilon_1^2) < \infty$ and a uniformly bounded second derivative of F. The Corollary 7.2.2 is an improvement of Boldin's result in the sense that F needs to be smooth only up to the first derivative and the r.v.'s need not have finite second moment.

Again, if $\mathbf{Y}_0$ and $\{\epsilon_i\}$ are so chosen that the Ergodic Theorem is applicable and $E(\mathbf{Y}_0) = \mathbf{0}$, then the coefficient $n^{-1}\Sigma_i \mathbf{Y}_{i-1}$ of the linear term in (14) will converge to $\mathbf{0}$, a.s.. Thus (14) becomes

$$\text{(14*)} \qquad \sup_{\|\mathbf{u}\|\leq B} |n^{1/2}\{F_n(x, \rho + n^{-1/2}\mathbf{u}) - F_n(x, \rho)\}| = o_p(1).$$

In particular, this implies that if $\hat{\rho}$ is an estimator of ρ such that

$$\|n^{1/2}(\hat{\rho} - \rho)\| = O_p(1),$$

then

$$\|n^{1/2}\{F_n(\cdot, \hat{\rho}) - F_n(\cdot, \rho)\}\|_\infty = o_p(1).$$

Consequently, *the estimation of* ρ *has asymptotically negligible effect on the estimation of the error d.f.* F. This is similar to the fact, observed in the previous chapter, that the estimation of the slope parameters in linear regression has asymptotically negligible effect on the estimation of the error d.f. as long as the design matrix is centered at the origin. □

An important application of (14) occurs when proving the a.u.l property of the serial rank correlations of the residuals as functions of $\mathbf{t}$. More precisely, let $R_{i\mathbf{t}}$ denote the rank of $X_i - \mathbf{t}'\mathbf{Y}_{i-1}$ among $X_j - \mathbf{t}'\mathbf{Y}_{j-1}$, $1 \leq j \leq n$, $1 \leq i \leq n$. Define $R_{i\mathbf{t}} = 0$ for $i \leq 0$. Rank correlations of lag j, for $1 \leq j \leq p$, are defined as

$$\text{(15)} \qquad S_j(\mathbf{t}) := \frac{12}{n(n^2-1)} \sum_{i=j+1}^{n} (R_{i-j\mathbf{t}} - \frac{(n+1)}{2})(R_{i\mathbf{t}} - \frac{(n+1)}{2}), \qquad \mathbf{t} \in \mathbb{R}^p,$$

$$\mathbf{S}' := (S_1, \ldots, S_p).$$

Simple algebra shows that

$$S_j(\mathbf{t}) = a_n[L_j(\mathbf{t}) - n(n+1)^2/4] + b_{nj}(\mathbf{t}), \qquad 1 \leq j \leq p,$$

where a_n is a nonrandom sequence not depending on $\mathbf{t}$, $|a_n| = O(1)$,

$$b_{nj}(\mathbf{t}) = \frac{6(n+1)}{\{n(n^2-1)\}}\Big(\sum_{i=n-j+1}^{n} + \sum_{i=1}^{j}\Big) R_{i\mathbf{t}},$$

and

$$L_j(\mathbf{t}) := n^{-3}\sum_{i=j+1}^{n} R_{i-j\mathbf{t}}\, R_{i\mathbf{t}}, \qquad 1 \le j \le p,\ \ \mathbf{t} \in \mathbb{R}^p.$$

Observe that $\sup\{|b_{nj}(\mathbf{t})|;\ \mathbf{t}\in\mathbb{R}^p\} \le 48p/n$, so that $n^{1/2}\sup\{|b_{nj}(\mathbf{t})|;\ \mathbf{t}\in\mathbb{R}^p\}$ tends to zero, a.s. It thus suffices to prove the a.u.l. of $\{L_j\}$ only, $1 \le j \le p$.

In order to state the a.u.l. result we need to introduce

$$\begin{aligned}
(16)\qquad Z_{ij} &:= f(\epsilon_{i-j})F(\epsilon_i) + f(\epsilon_i)F(\epsilon_{i-j}), && i > j,\\
&:= 0, && i \le j.\\
\mathbf{U}_{ij} &:= \mathbf{Y}_{i-j-1}\, F(\epsilon_i)f(\epsilon_{i-j}) + \mathbf{Y}_{i-1}\, f(\epsilon_i)F(\epsilon_{i-j}), && i > j,\\
&:= 0, && i \le j.\\
\bar{Z}_j &:= n^{-1}\sum_{i=j+1}^{n} Z_{ij}, \qquad \bar{\mathbf{U}}_j := n^{-1}\sum_{i=j+1}^{n}\mathbf{U}_{ij}, && 1 \le j \le p.\\
\bar{\mathbf{Y}}_n &:= n^{-1}\sum_{i=1}^{n}\mathbf{Y}_{i-1}.
\end{aligned}$$

Observe that $\{Z_{ij}\}$ are bounded r.v.'s with $EZ_{ij} = \int f^2(x)\,dx$ for all i and j. Moreover, $\{\epsilon_i\}$ i.i.d. F imply that $\{Z_{ij},\ j<i\le n\}$ are stationary and ergodic. By the Ergodic Theorem

$$\bar{Z}_j \longrightarrow b(f) := \int f^2(x)\,dx, \quad \text{a.s.}, \qquad j = 1, \dots, p.$$

We are now ready to state and prove

Theorem 7.2.2. *Assume that* (7.1.1), (a2), (a3*) *and* (a4) *hold. Then, for every* $0 < B < \infty$ *and for every* $1 \le j \le p$,

$$(17)\qquad \sup_{\|\mathbf{u}\|\le B}\big|n^{1/2}[L_j(\rho+n^{-1/2}\mathbf{u}) - L_j(\rho)] - \mathbf{u}'[b(f)\bar{\mathbf{Y}}_n - \bar{\mathbf{U}}_j]\big| = o_p(1).$$

If (a2) *and* (a3*) *are strengthened to requiring* $E(\|\mathbf{Y}_0\|^2 + \epsilon_1^2) < \infty$ *and* $\{X_i\}$ *stationary and ergodic then* $\bar{\mathbf{Y}}_n$ *and* $\bar{\mathbf{U}}_j$ *may be replaced by their respective expectations in* (17).

Proof. Fix a j in $1\le j\le p$. For the sake of simplicity of the exposition, write $L(\mathbf{u})$, $L(0)$ for $L_j(\rho+n^{-1/2}\mathbf{u})$, $L_j(\rho)$, respectively. Apply similar convention to other functions of $\mathbf{u}$. Also write $\epsilon_{i\mathbf{u}}$ for $\epsilon_i - n^{-1/2}\mathbf{u}'\mathbf{Y}_{i-1}$ and $F_n(\cdot)$ for $F_n(\cdot, \rho)$. With these conventions $R_{i\mathbf{u}}$ is now the rank of

$X_i - (\rho + n^{-1/2}\mathbf{u})'\mathbf{Y}_{i-1} = \epsilon_{i\mathbf{u}}$. In other words, $R_{i\mathbf{u}} \equiv nF_n(\epsilon_{i\mathbf{u}}, \mathbf{u})$ and

$$L(\mathbf{u}) = n^{-1}\sum_{i=j+1}^{n} F_n(\epsilon_{i-j\mathbf{u}}, \mathbf{u})\, F_n(\epsilon_{i\mathbf{u}}, \mathbf{u}), \qquad \mathbf{u} \in \mathbb{R}^p.$$

The proof is based on the linearity properties of $F_n(\cdot, \mathbf{u})$ as given in (14) of Corollary 7.2.2 above. In fact if we let

$$B_n(x, \mathbf{u}) := F_n(x, \mathbf{u}) - F_n(x) - n^{-1/2}\mathbf{u}'\mathbf{Y}_n f(x), \qquad x \in \mathbb{R}.$$

then (14) is equivalent to

$$\sup n^{1/2}|B_n(x, \mathbf{u})| = o_p(1).$$

All supremums, unless specified otherwise, in the proof are over $x \in \mathbb{R}$, $1 \le i \le n$ and/or $\|\mathbf{u}\| \le B$. Rewrite

$$\begin{aligned} n^{1/2}&(L(\mathbf{u}) - L(0)) \\ &= n^{-1/2}\sum_{i=j+1}^{n}\{F_n(\epsilon_{i-j\mathbf{u}}, \mathbf{u})\, F_n(\epsilon_{i\mathbf{u}}, \mathbf{u}) - F_n(\epsilon_{i-j})\, F_n(\epsilon_i)\} \\ &= n^{-1/2}\sum_{i=j+1}^{n}[\{B_n(\epsilon_{i-j\mathbf{u}}, \mathbf{u}) + F_n(\epsilon_{i-j\mathbf{u}}) + n^{-1/2}\mathbf{u}'\mathbf{Y}_n f(\epsilon_{i-j\mathbf{u}})\}\cdot \\ &\qquad\qquad \cdot\{B_n(\epsilon_{i\mathbf{u}}, \mathbf{u}) + F_n(\epsilon_{i\mathbf{u}}) + n^{-1/2}\mathbf{u}'\mathbf{Y}_n f(\epsilon_{i\mathbf{u}})\} \\ &\qquad\qquad\qquad - F_n(\epsilon_{i-j})\, F_n(\epsilon_i)]. \end{aligned}$$

Hence, from (14), (a2) and (a3*),

$$\begin{aligned} (18)\quad n^{1/2}&(L(\mathbf{u}) - L(0)) \\ &= n^{-1/2}\sum_{i=j+1}^{n}[F_n(\epsilon_{i-j\mathbf{u}})\, F_n(\epsilon_{i\mathbf{u}}) - F_n(\epsilon_i)\, F_n(\epsilon_{i-j})] \\ &\quad + n^{-1}\sum_{i=j+1}^{n}[F_n(\epsilon_{i-j\mathbf{u}}) f(\epsilon_{i\mathbf{u}}) + F_n(\epsilon_{i\mathbf{u}}) f(\epsilon_{i-j\mathbf{u}})](\mathbf{u}'\mathbf{Y}_n) + \bar{o}_p(1), \end{aligned}$$

where, now, $\bar{o}_p(1)$ is a sequence of stochastic processes converging to zero uniformly, in probability, over the set $\mathcal{N}(B)$.

Now recall that (a4) and the asymptotic uniform continuity of the standard empirical process based on i.i.d. r.v.'s imply that

$$\sup_{|x-y| \le \delta} n^{1/2}|[F_n(x) - F(x)] - [F_n(y) - F(y)]| = o_p(1)$$

when first $n \to \infty$ and then $\delta \to 0$. Hence from (a2) and the fact that

$$\sup_{i,u} |\epsilon_{iu} - \epsilon_i| \le B n^{-1/2} \max_i \|Y_{i-1}\|,$$

one readily obtains

$$\sup_{i,u} n^{1/2} |[F_n(\epsilon_{iu}) - F(\epsilon_{iu})] - [F_n(\epsilon_i) - F(\epsilon_i)]| = o_p(1).$$

From this and (a4) we obtain

$$\text{(19)} \qquad \sup_{i,u} n^{1/2} |F_n(\epsilon_{iu}) - F_n(\epsilon_i) + n^{-1/2} u' Y_{i-1} f(\epsilon_i)| = o_p(1).$$

From (18), (19), the uniform continuity of f and F, the Glivenko–Cantelli lemma, one obtains

$$\begin{aligned}\text{(20)} \quad & n^{1/2}(L(u) - L(0)) \\ &= n^{-1} \sum_{i=j+1}^{n} [F(\epsilon_{i-j}) f(\epsilon_i) + F(\epsilon_i) f(\epsilon_{i-j})](u' Y_n) \\ &\qquad - u' n^{-1} \sum_{i=j+1}^{n} \{Y_{i-j-1} f(\epsilon_{i-j}) F(\epsilon_i) + Y_{i-1} f(\epsilon_i) F(\epsilon_{i-j})\} + \bar{o}_p(1).\end{aligned}$$

In concluding (20) we also used the fact that by (a2) and (a3*),

$$\sup_u |n^{-3/2} \sum_{i=j+1}^{n} |u' Y_{i-j} \cdot u' Y_{i-1}| \le B\, n^{-1/2} \max_i \|Y_{i-1}\|\, n^{-1} \sum_{i=j+1}^{n} \|Y_{i-j}\| = o_p(1).$$

Now (17) readily follows from (20) and the notation introduced just before the statement of the theorem. The rest is obvious. □

Remark 7.2.4. *Autoregressive moving average models.* Boldin (1989) and Kreiss (1991) give an analogue of (14*) for a moving average model of order q and an autoregressive moving average model of order (p,q) (ARMA(p,q)), respectively, when the error d.f. F has zero mean, finite second moment and bounded second derivative. Here we shall illustrate as to how Theorem 2.2b.1 can be used to yield the same result under weaker conditions on F. For the sake of clarity, the details are carried out for an ARMA(1,1) model only.

Let $\epsilon_0, \epsilon_1, \epsilon_2, \ldots,$ be i.i.d. F r.v.'s and X_0 be a r.v. independent of $\{\epsilon_i, i \ge 1\}$. Consider the process given by the relation

$$\text{(21)} \qquad X_i = \rho X_{i-1} + \epsilon_i + \beta\, \epsilon_{i-1}, \qquad i \ge 1,$$

where $|\rho| < 1$, $|\beta| < 1$. One can rewrite this model as

$$\begin{aligned}\text{(22)} \qquad \epsilon_i &= X_1 - (\rho X_0 + \beta \epsilon_0), & i = 1, \\ &= X_i - \sum_{j=1}^{i-1} (-\beta)^j (\rho+\beta)\, X_{i-j-1} + (-\beta)^{i-1} (\rho X_0 + \beta \epsilon_0), & i \ge 2.\end{aligned}$$

Let $\boldsymbol{\theta} := (s, t)'$ denote a point in the open square $(-1, 1)^2$ and $\boldsymbol{\theta}_0 := (\rho, \beta)'$ denote the true parameter value. Assume that $\boldsymbol{\theta}$'s are restricted to the following sequence of neighborhoods: For a $b \in (0, \infty)$,

$$n^{1/2}\{|s-\rho| + |t-\beta|\} \leq b. \tag{23}$$

Let $\{\tilde{\epsilon}_i, i \geq 1\}$ stand for the residuals $\{\epsilon_i, i \geq 1\}$ of (22) after ρ and β are replaced by s and t, respectively, in (22). Let $F_n(\cdot, \boldsymbol{\theta})$ denote the empirical process of $\{\tilde{\epsilon}_i, 1 \leq i \leq n\}$. This empirical can be rewritten as

$$F_n(x, \boldsymbol{\theta}) = n^{-1}\sum_{i=1}^{n} I(\epsilon_i \leq x + \delta_{ni}), \qquad x \in \mathbb{R}, \tag{24}$$

where

$$\begin{aligned}
\delta_{ni} &:= (s-\rho)X_0 + (t-b)\,\epsilon_0, & i = 1, \\
&= \sum_{j=1}^{i-2}[(-t)^j(s+t) - (-\beta)^j(\rho+\beta)]\, X_{i-j-1} & \\
&\qquad + (-t)^{i-1}(sX_0 + t\epsilon_0) - (-\beta)^{i-1}(\rho X_0 + \beta\epsilon_0), & i \geq 2. \\
&= \delta_{ni1} + \delta_{ni2}, \qquad \text{say}, & i \geq 2.
\end{aligned} \tag{25}$$

From (25), it follows that for every $\boldsymbol{\theta} \in (-1, 1)^2$ satisfying (23),

$$|\delta_{n1}| \leq bn^{-1/2}(|X_0| + |\epsilon_0|),$$

$$\max_{2 \leq i \leq n} |\delta_{ni1}| \leq 2b\, n^{-1/2} \max_{1 \leq i \leq n} |X_i|\, (1 - bn^{-1/2} - \beta)^{-1}\{1 + (1-|\beta|)^{-1}\},$$

$$\max_{2 \leq i \leq n} |\delta_{ni2}| \leq 2bn^{-1/2}(1 - bn^{-1/2} - \beta)^{-1}(|X_0| + |\epsilon_0|).$$

Consequently, if $n^{-1/2}\max_{1 \leq i \leq n} |X_i| = o_p(1)$, then the $\{\delta_{ni}\}$ of (25) would satisfy (2.2b.A2) for every $\boldsymbol{\theta} \in (-1, 1)^2$. But by (21),

$$\begin{aligned}
X_i &= \rho X_0 + \beta\epsilon_0 + \epsilon_1, & i = 1, \\
&= \rho^{i-1}(\rho X_0 + \beta\epsilon_0) + \sum_{j=1}^{i-2} \rho^j(\rho+\beta)\, \epsilon_{i-j-1} + \epsilon_i, & i \geq 2.
\end{aligned} \tag{26}$$

Therefore, (2.2b.A2) will hold for the above $\{\delta_{ni}\}$ if

$$n^{-1/2}\max_{1 \leq i \leq n} |\epsilon_i| = o_p(1). \tag{27}$$

We now verify (2.2b.A3) for the above $\{\delta_{ni}\}$ and with $h_{ni} \equiv 1$. That is we must show that $n^{-1/2} \sum_i |\delta_{ni}| = O_p(1)$. We proceed as follows. Let $u = n^{1/2}(s - \rho)$, $v = n^{1/2}(t - \beta)$ and $Z_0 := |X_0| + |\epsilon_0|$. By (23), $|u| + |v| \le b$. From (25),

$$
\begin{aligned}
(28)\qquad n^{-1/2} \sum_i |\delta_{ni}| &\le n^{-1} b Z_0 \\
&\quad + n^{-1/2} \sum_{i=2}^{n} \Big| \sum_{j=0}^{i-2} [(-t)^j (s+t) - (-\beta)^j (\rho+\beta)]\, X_{i-j-1} \Big| \\
&\quad + n^{-1/2} \sum_{i=2}^{n} |(-t)^{i-1}(sX_0 + t\epsilon_0) - (-\beta)^{i-1} Z|, \\
&= A_{n1} + A_{n2} + A_{n3}, \quad \text{say.}
\end{aligned}
$$

Clearly, $|A_{n1}| = o(1)$, a.s. Rewrite

$$
\begin{aligned}
A_{n2} &= n^{-1/2} \sum_{i=2}^{n} \Big| \sum_{j=0}^{i-2} [(-t)^j \{(u+v)n^{-1/2} + \rho + \beta\} - (-\beta)^j (\rho+\beta)]\, X_{i-j-1} \Big| \\
&\le 2bn^{-1} \sum_{i=2}^{n} \sum_{j=0}^{i-2} |t|^j |X_{i-j-1}| + 2\, n^{-1/2} \sum_{i=2}^{n} \Big| \sum_{j=0}^{i-2} [(-t)^j - (-\beta)^j] X_{i-j-1} \Big| \\
(29)\qquad &= 2bA_{n21} + 2A_{n22}, \quad \text{say.}
\end{aligned}
$$

By a change of variables and an interchange of summations one obtains

$$
(30)\qquad A_{n21} \le n^{-1} \sum_{i=1}^{n} |X_i| \, (1 - |t|)^{-1}.
$$

Next, use the expansion $a^j - c^j = (a - c) \sum_{k=0}^{j-1} a^{j-1-k} c^k$ for any real numbers a, c, to obtain

$$
A_{n22} \le b\, n^{-1} \sum_{i=3}^{n} \sum_{j=1}^{i-2} \sum_{k=0}^{j-1} |t|^{j-1-k} \, |\beta|^k |X_{i-j-1}|.
$$

Again, use change of variables and interchange of summations repeatedly and the fact that $|\beta| \vee |t| < 1$, to conclude that this upper bound is bounded above by

$$
b(1 - |\beta|)^{-1} [(1 - |t|)^{-1} + 1]\, n^{-1} \sum_{i=1}^{n} |X_i|.
$$

This, (28) and (29) together with (23) imply that

(31) $$A_{n2} \le 2b\, n^{-1} \sum_{i=1}^{n} |X_i|\, [(1-bn^{-1/2}-|\beta|)^{-1}\{1+(1-|\beta|)^{-1}\} + (1-|\beta|)^{-1}].$$

Finally, similar calculations show that

(32) $$A_{n3} = O_p(n^{-1/2}).$$

From (28), (31) and (32) it thus follows that if $n^{-1}\sum_{i=1}^{n}|X_i| = O_p(1)$, then the $\{\delta_{ni}\}$ of (25) will satisfy (2.2b.A3) with $h_{ni} \equiv 1$. But in view of (26) and the assumption that $|\rho| \vee |\beta| < 1$, it readily follows that if

(33) $$n^{-1}\sum_{i=1}^{n}|\epsilon_i| = O_p(1),$$

then (2.2b.A3) with $h_{ni} \equiv 1$ holds for the $\{\delta_{ni}\}$ of (25). We have thus proved the following:

If (21) holds with the error d.f. F satisfying (**F1**), (**F2**), (27) and (33), then $\forall\ \boldsymbol{\theta} \in (-1, 1)^2$,

$$\sup_x |n^{-1/2} \sum_{i=1}^{n} \{I(\tilde{\epsilon}_i \le x) - I(\epsilon_i \le x) - F(x+\delta_{ni}) + F(x)\}| = o_p(1).$$

Now use an argument like the one used in the proof of Theorem 7.2.1 to conclude the following

Corollary 7.2.3. *In addition to* (21), *assume that the error d.f.* F *satisfies* (**F1**), (**F2**), (27) *and* (33). *Then,* $\forall\ 0 < b < \infty$,

$$\sup |n^{1/2}[F_n(x,\boldsymbol{\theta}) - F_n(x,\boldsymbol{\theta}_0)] - n^{-1/2}\, \Sigma_i\, \delta_{ni}\, f(x)| = o_p(1),$$

where the supremum is taken over $x \in \mathbb{R}$ *and* $\boldsymbol{\theta}$, $\boldsymbol{\theta}_0$ *satisfying* (23).

If (33) *is strengthened to assuming that* $E|\epsilon| < \infty$, *then*

$$\sup |n^{-1/2} \sum_i \delta_{ni} - n^{1/2}[(s-\rho)(1-\rho)^{-1} + (t-\beta)(1+\beta)^{-1}]\, \mu| = o_p(1),$$

where the supremum is taken over s, t *satisfying* (23) and $\mu = E\epsilon$. □

Consequently, if $E\epsilon = 0$ and $(\hat{\rho}, \hat{\beta})$ is an estimator of (ρ, β) such that $\|n^{1/2}(\hat{\rho}-\rho,\ \hat{\beta}-\beta)\| = O_p(1)$, then an analogue of (14*) holds in the present case also under weaker conditions than those given by Boldin or Kreiss.

The details for proving an anlogue of Corollary 7.2.3 for a general ARMA(p,q) model are similar but some what complicated to those given above. □

7.3. GM- and R- Estimators.

In this section we shall discuss the asymptotic distributions of GM- and R-estimators of ρ. In addition, some consistent estimators of the functional Q(f) will be also constructed. We begin with

7.3a. GM-Estimators.

Here we shall state the asymptotic normality of the GM-estimators. Let $\rho_{\mathbf{M}}$ stand for a solution of (7.1.3) such that $\|n^{1/2}(\rho_{\mathbf{M}} - \rho)\| = O_p(1)$. That such an estimator $\rho_{\mathbf{M}}$ exists can be seen by an argument similar to the one given in Huber (1981) in connection with the linear regression model. To state the asymptotic normality of $\rho_{\mathbf{M}}$ we need to introduce some more notation. Let

$$(1)\qquad \mathscr{X} := \begin{bmatrix} X_0 & X_{-1},\dots,X_{1-p} \\ X_1 & X_0,\dots\dots,X_{2-p} \\ \vdots & \vdots \qquad \vdots \\ X_{n-1} & X_{n-2},\dots,X_{n-p} \end{bmatrix}, \quad \mathscr{G} := \begin{bmatrix} g(X_0) & g(X_{-1}),\dots,g(X_{1-p}) \\ g(X_1) & g(X_0),\dots,g(X_{2-p}) \\ \vdots & \vdots \qquad \vdots \\ g(X_{n-1}) & g(X_{n-2}),\dots,g(X_{n-p}) \end{bmatrix},$$

$$\mathcal{G}' := n^{1/2}(\mathcal{G}_1(\rho),\dots,\mathcal{G}_p(\rho)), \quad \mathcal{B}_n := \mathscr{G}'\mathscr{X} = \Sigma_i\,(g(X_{i-1})Y'_{i-1},\dots,g(X_{i-p})Y'_{i-1})'.$$

Proposition 7.3.1. *In addition to* (7.1.1), (7.2.a2), (7.2.a4), (7.2.b1), (7.2.b2) and (7.2.b3) *assume that*

(b4) $\quad n^{-1}\mathcal{B}_n = \mathcal{B} + o_p(1)$, *for some* $p\times p$ *non-random positive definite matrix* $\mathcal{B}$.

Then

$$n^{1/2}(\rho_{\mathbf{M}} - \rho) = -(\mathcal{B}a)^{-1}\mathcal{G} + o_p(1).$$

If, in addition, we assumes that

(b5) $\quad n^{-1}\mathscr{G}'\mathscr{G} = G^* + o_p(1)$, G^* *a* $p\times p$ *non-random positive definite matrix,*

then

$$n^{1/2}(\rho_{\mathbb{M}} - \rho) \underset{d}{\rightarrow} N(0, \mathbf{J}\,), \qquad \mathbf{J} := (\int f\, d\psi)^{-2} \cdot (\int \psi^2\, dF) \cdot \mathcal{B}^{-1} G^* \mathcal{B}^{-1}.$$

Proof. Follows from Corollary 7.2.1, the Cramer–Wold device and Lemma A.3 in the appendix applied to $\mathcal{G}$. □

Again, if Y_0 and $\{\epsilon_i\}$ are so chosen as to make $\{X_i\}$ stationary, ergodic and $E(\|Y_0\|^2 + \epsilon_1^2) < \infty$ then (b4) and (b5) are *a priori* satisfied. See, e.g., Anderson (1971; p 203).

Note: For a more general class of GM—estimators see Bustos (1982) where a result analogous to the above corollary for smooth score functions ψ is obtained. □

7.3b. R-Estimators.

This section will discuss an analogue of Jaeckel's (1972) R-estimators of ρ and their large sample properties.

Recall that R_{it} is the rank of $X_i - \mathbf{t}'Y_{i-1}$ among $\{X_k - \mathbf{t}'Y_{k-1},\ 1 \le k \le n\}$, for $1 \le i \le n$. Also, $R_{it} \equiv 0$ for $i \le 0$. Let φ be a nondecreasing score function from $[0, 1]$ to the real line such that

$$\sum_{i=1}^{n} \varphi(i/(n+1)) = 0. \tag{1}$$

For example, if $\varphi(t) = -\varphi(1-t)$ for all $t \in [0.\ 1]$, i.e., if φ is skew symmetric, then it satisifies (1). Define

$$\mathcal{S}_j(\mathbf{u}) := n^{-1} \sum_{i=j+1}^{n} X_{i-j}\ \varphi(R_{i\mathbf{u}}/(n+1)), \qquad 1 \le j \le p,\ \mathbf{u} \in \mathbb{R}^p,$$
$$\mathcal{S}' := (\mathcal{S}_1, \ldots, \mathcal{S}_p).$$

The class of rank statistics $\mathcal{S}$, one for each φ, is an analogue of the class of rank statistics discussed in Section 4.3 above in connection with linear regression models where one replaces the weights $\{X_{i-j}\}$ by appropriate design points. A test of the hypothesis $\rho = \rho_0$ may be based on a suitably standardized $\mathcal{S}(\rho_0)$, the large values of the statistic being significant.

It is thus natural to define R-estimators of ρ by the relationship

$$\tilde{\rho}_R = \arg\min\{\|\mathcal{S}(\mathbf{t})\|;\ \mathbf{t} \in \mathbb{R}^p\}. \tag{2}$$

An alternative way to define R-estimators of ρ is to adapt Jaeckel (1972) to the AR(p) situation. Accordingly, for a $\mathbf{t} \in \mathbb{R}^p$, let

$$Z_k(\mathbf{t}) = X_k - \mathbf{t}'Y_{k-1},\ 1 \le k \le n,$$

$$Z_{(i)}(\mathbf{t}) := \text{the ith largest residual among } \{Z_k(\mathbf{t}),\ 1 \le k \le n\},\ 1 \le i \le n,$$

$$\mathscr{J}(\mathbf{t}) := \sum_{i=1}^{n} \varphi(i/(n+1))\, Z_{(i)}(\mathbf{t}) \equiv \sum_{i=1}^{n} \varphi(R_{it}/(n+1))(X_i - \mathbf{t}'\mathbf{Y}_{i-1}).$$

Then Jaeckel's estimator $\tilde{\rho}_J$ is defined by the relation

$$\tilde{\rho}_J = \arg\min\{\mathscr{J}(\mathbf{t});\ \mathbf{t} \in \mathbb{R}^p\}.$$

Jaeckel's argument about the existence of an analogue of $\tilde{\rho}_J$ in the context of linear regression model can be adapted to the present situation. This follows from the following *three* lemmas, the first of which is of a general interest.

Lemma 7.3b.1. *Let* $d_1, d_2, \ldots, d_n, v_1, v_2, \ldots, v_n$, *be real numbers such that not all* $\{d_i\}$ *are the same and no two* $\{v_i\}$ *are the same. Let* r_{iu} *denote the rank of* $v_i - ud_i$ *among* $\{v_j - ud_j;\ 1 \le j \le n\}$, $u \in \mathbb{R}$. *Let* $\{b_n(i);\ 1 \le i \le n\}$ *be a set of real numbers that are nondecreasing in* i. *Let*

$$T(u) := \sum_{i=1}^{n} d_i\, b_n(r_{iu}), \qquad u \in \mathbb{R}.$$

Then, $T(u)$ *is a nonincreasing step function in all those* $u \in \mathbb{R}$ *for which there are no ties among* $\{v_j - ud_j;\ 1 \le j \le n\}$.

Proof. See Theorem II.7E, p35 of Hájek (1969). □

Lemma 7.3b.2. *Assume that the model* (7.1.1) *holds with* $(\mathbf{Y}_0', X_1, X_2, \ldots, X_n)$ *having a continuous joint distribution. Then the following hold.*

(a) *For each realization* $(\mathbf{Y}_0', X_1, X_2, \ldots, X_n)$, *the assumption* (1) *implies that* $\mathscr{J}(\mathbf{t})$ *is nonnegative, continuous and convex function of* $\mathbf{t}$ *with its a.e. derivative equal to* $-n\mathcal{S}(\mathbf{t})$.

(b) *If the realization* $(\mathbf{Y}_0', X_1, X_2, \ldots, X_n)$ *is such that the rank of* $\mathscr{X}_c$ *is* p *then, for every* $0 < b < \infty$, *the set* $\{\mathbf{t} \in \mathbb{R}^p;\ \mathscr{J}(\mathbf{t}) \le b\}$ *is bounded, where* $\mathscr{X}_c$ *is the* $\mathscr{X}$ *of* (7.3a.1), *centered at the origin.*

Proof. (a). For any $\mathbf{x}' = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$, let $x(1) \le x(2) \le \ldots \le x(n)$ denote the ordered $x_1, x_2, \ldots, x_n$. Let $\Pi := \{\boldsymbol{\pi} = (\pi_1, \pi_2, \ldots, \pi_n)';\ \boldsymbol{\pi}$ a permutation of the integers $1, 2, \ldots, n.\}$, $b_n(i) := \varphi(i/(n+1))$, $1 \le i \le n$, and define

$$D(\mathbf{x}) := \sum_{i=1}^{n} b_n(i)\, x(i), \qquad D_{\boldsymbol{\pi}}(\mathbf{x}) := \sum_{i=1}^{n} b_n(i)\, x_{\pi_i}, \qquad \mathbf{x} \in \mathbb{R}^n,$$

$$k := \min\{1 \le j \le n;\ b_n(j) > 0\}.$$

Observe that $\mathscr{J}(\mathbf{t}) = D(\mathbf{Z}(\mathbf{t}))$.

Now, (1) and φ nondecreasing implies that

$$\begin{aligned} D(\mathbf{x}) &= \sum_{i=1}^{n} b_n(i)\,(x(i) - x(k)) \\ &= \sum_{i=1}^{k-1} b_n(i)\,(x(i) - x(k)) + \sum_{i=k}^{n} b_n(i)\,(x(i) - x(k)) \\ &\geq 0, \qquad\qquad \forall\ \mathbf{x} \in \mathbb{R}^n, \end{aligned}$$

because each summand is nonnegative. This proves that $\mathcal{J}(\mathbf{t}) \geq 0$, $\mathbf{t} \in \mathbb{R}$.

By Theorem 368 of Hardy, Littlewood and Polya (1952),

$$D(\mathbf{x}) = \max_{\pi\in\Pi} D_{\pi}(\mathbf{x}), \qquad \forall\ \mathbf{x} \in \mathbb{R}^n.$$

Therefore, $\forall\ \mathbf{t} \in \mathbb{R}^p$,

$$(*) \qquad \begin{aligned} \mathcal{J}(\mathbf{t}) = D(Z(\mathbf{t})) &= \max_{\pi\in\Pi} D_{\pi}(Z(\mathbf{t})) \\ &= \max_{\pi\in\Pi} \sum_{i=1}^{n} b_n(i)(X_{\pi_i} - \mathbf{t}'\mathbf{Y}_{\pi_{i-1}}). \end{aligned}$$

This shows that $\mathcal{J}(\mathbf{t})$ is a maximal element of a finite number of continuous and convex functions, which itself is continuous and convex. The statement about a.e. differential being $-n\mathcal{S}(\mathbf{t})$ is obvious. This completes the proof of (a).

(b) Without the loss of generality assume $b > \mathcal{J}(\mathbf{0})$. Write a $\mathbf{t}\in\mathbb{R}^p$ as $\mathbf{t} = u\boldsymbol{\theta}$, $u\in\mathbb{R}$, $\boldsymbol{\theta}\in\mathbb{R}^p$, $\|\boldsymbol{\theta}\| = 1$. Let $d_i \equiv \boldsymbol{\theta}'\mathbf{Y}_{i-1}$. The assumptions about $\mathcal{X}$ imply that not all $\{d_i\}$ are equal. Rewrite

$$\mathcal{J}(\mathbf{t}) = \mathcal{J}(u\boldsymbol{\theta}) = \sum_{i=1}^{n} b_n(i)\,(X - ud)(i) = \sum_{i=1}^{n} b_n(r_{iu})(X_i - ud_i)$$

where now r_{iu} is the rank of $X_i - ud_i$ among $\{X_j - ud_j;\ 1 \leq j \leq n\}$. From (*) above, it follows that $\mathcal{J}(u\boldsymbol{\theta})$ is linear and convex in u, for every $\boldsymbol{\theta}\in\mathbb{R}^p$, $\|\boldsymbol{\theta}\| = 1$. Its a.e. derivative w.r.t. u is $-\sum_{i=1}^{n} d_i b_n(r_{iu})$, which by Lemma 7.3b.1 and because of the assumed continuity, is nondecreasing in u and eventually positive. Hence $\mathcal{J}(u\boldsymbol{\theta})$ will eventually exceed b, for every $\boldsymbol{\theta}\in\mathbb{R}^p$, $\|\boldsymbol{\theta}\| = 1$.

Thus, there exists a $u_{\boldsymbol{\theta}}$ such that $\mathcal{J}(u_{\boldsymbol{\theta}}\boldsymbol{\theta}) > b$. Since $\mathcal{J}$ is continuous, there is an open set $O_{\boldsymbol{\theta}}$ of unit vectors $\boldsymbol{\nu}$, containing $\boldsymbol{\theta}$ such that $\mathcal{J}(u_{\boldsymbol{\theta}}\boldsymbol{\nu}) > b$. Since $b > \mathcal{J}(\mathbf{0})$, and $\mathcal{J}$ is convex, $J(u\boldsymbol{\nu}) > b$, $\forall\ u \geq u_{\boldsymbol{\theta}}$ and $\forall\ \boldsymbol{\nu}\in O_{\boldsymbol{\theta}}$. Now, for each unit vector $\boldsymbol{\theta}$, there is an open set $O_{\boldsymbol{\theta}}$ covering it. Since the unit sphere is compact, a finite number of these sets covers it. Let m be the maximum of the corresponding finite set of $u_{\boldsymbol{\theta}}$. Then for all $u \geq m$, for all unit vectors $\boldsymbol{\nu}$, $\mathcal{J}(u\boldsymbol{\nu}) > b$. This proves the claim (b) and also the lemma. □

Note: Lemma 7.3b.2 *and its proof is an adaptation of Theorems 1 and 2 of Jaeckel* (1972) *to the present case.* □

From the above lemma it follows that if the r.v.'s $Y_0, X_1, X_2, \ldots, X_n$ are continuous and the matrix $n^{-1}\Sigma_i(Y_{i-1} - \bar{Y})(Y_{i-1} - \bar{Y})'$ is a.s. positive definite, then the rank of $\mathcal{K}$ is a.s. p and the set $\{t \in \mathbb{R}^p;\ \mathcal{J}(t) \le b\}$ is a.s. bounded for every $0 \le b < \infty$. Thus a minimizer $\tilde{\rho}_J$ of $\mathcal{J}$ exists a.s. and has the property that makes $\|\mathcal{S}\|$ small. As is shown in Jaeckel (1972) in connection with the linear regression model, it will follow from the linearity result given in Theorem 7.3b.1 below that $\tilde{\rho}_J$ and $\tilde{\rho}_R$ are asymptotically equivalent. Note that the score function φ need not satisfy (1) in this theorem.

Some steps of the proof of Theorem 7.3b.1 heavily depend on the representation of the AR(p) process $\{X_i\}$ in terms of the error variables $\{\epsilon_i\}$. For that reason we shall now extend the index i in the process $\{X_i\}$ to both sides of 0. Accordingly, assume that $\{\epsilon_i,\ i = 0, \pm1, \pm2, \ldots.\}$ are i.i.d. F r.v.'s and that

(3) $$X_i = \rho_1 X_{i-1} + \rho_2 X_{i-2} + \ldots + \rho_p X_{i-p} + \epsilon_i, \qquad i = 0, \pm1, \pm2, \ldots\ , \quad \rho \in \mathbb{R}^p.$$

In addition assume the following:

(4) *All roots of the equation*

$$x^p - \rho_1 x^{p-1} - \rho_2 x^{p-2} - \ldots - \rho_p = 0 \ \text{ are in the interval } (-1, 1).$$

It is well known that if $E|\epsilon|^2 < \infty$, there exist constants $\{\theta_j,\ j \ge 0\}$ such that $\theta_0 = 1$, $\Sigma_{j\ge0} |\theta_j| < \infty$, and that

(5) $$X_i = \sum_{k\le i} \theta_{i-k}\, \epsilon_k, \qquad i = 0, \pm1, \pm2, \ldots, \text{ in } L_2 \text{ and a.s.},$$

where the unspecified lower limit on the index of summation is $-\infty$. See, e.g., Anderson (1971) and Brockwell and Davis (1987, pp 76–86). Thus $\{X_i\}$ is stationary, ergodic and $E\|Y_0\|^2 < \infty$. Hence (7.2.a1) implies (7.2.a3). Moreover, the stationarity of $\{Y_{i-1}\}$ and $E\|Y_0\|^2 < \infty$ imply that $\forall\ \eta > 0$,

(6) $$P(\max_{1\le i\le n} \|Y_{i-1}\| \ge \eta\, n^{1/2}) \le \{\eta n^{1/2}\}^{-2} \sum_{i=1}^{n} E\|Y_{i-1}\|^2\, I(\|Y_{i-1}\| \ge \eta n^{1/2})$$

$$= \eta^{-2}\, E\|Y_0\|^2 I(\|Y_0\| \ge \eta n^{1/2}) = o(1).$$

Thus (7.2.a2) holds. These observations will be used in the sequel frequently, some times without mentioning.

With this preliminary background, we now state

Theorem 7.3b.1. (*A.U.L. of R–statistics*). *Assume that* (3) *and* (4) *above hold. In addition, assume that* F *satisfies* **(F1)**, **(F2)** *and that the following hold.*

(c1) (i) $E\epsilon = 0$. (ii) $0 < E\epsilon^4 < \infty$.

(c2) φ *is nondecreasing and differentiable with its derivative* $\dot{\varphi}$ *being uniformly continuous on* $[0, 1]$.

Then, for every $0 < B < \infty$,

$$\sup_{\|u\| \le B} \|n^{1/2}\{S(\rho + n^{-1/2}u) - \hat{S}\} + u'\Sigma Q\| = o_p(1), \tag{7}$$

where $\hat{S}' := (\hat{S}_1, \ldots, \hat{S}_p)$ *with*

$$\hat{S}_j := n^{-1} \sum_{i=j+1}^{n} (X_{i-j} - \bar{X}_j)[\varphi(F(\epsilon_i)) - \bar{\varphi}], \quad \bar{\varphi} = \int_0^1 \varphi(t)dt, \ 1 \le j \le p,$$

$$\bar{X}_j := n^{-1} \sum_{i=j+1}^{n} X_{i-j}, \qquad Q := \int f \, d\varphi(F),$$

$$\Sigma := ((\beta(k-j)), 1 \le k \le p; \ 1 \le i \le p; \ \beta(k) = \mathrm{Cov}(X_0, X_k), 1 \le k \le p.$$

Before proceeding to prove the above result, we shall state a lemma giving the asymptotic continuity of certain basic r.w.e.p.'s. Accordingly, let h be a nonnegative measurable function from [0, 1] to $\mathbb{R}$, U denote a uniform [0, 1] r.v., and define

$$\mathcal{Z}_j(t) := n^{-1/2} \sum_{i=1}^{n} X_{i-j}\,[h(F(\epsilon_i))\, I(F(\epsilon_i) \le t) - H(t)], \ 0 \le t \le 1, \ 1 \le j \le p,$$

where $H(t) := E\, h(U) I(U \le t) = \int_0^t h(s)ds, \ 0 \le t \le 1$.

The proof of the following lemma will be given in the subsection 7.3d.

Lemma 7.3b.3. *In addition to* (3), (4) *and* (c1(i)) *assume that* $Eh^4(U) < \infty$. *Then*, $\forall \ \gamma > 0, \ \forall \ 1 \le j \le p$,

$$\lim_{\eta \to 0} \limsup_n P\Big(\sup_{|t-v| \le \eta} |\mathcal{Z}_j(t) - \mathcal{Z}_j(v)| > \gamma\Big) = 0. \qquad \square$$

Proof of Theorem 7.3b.1. Observe that with

$$\tilde{S}_j(\mathbf{u}) := n^{-1}\sum_{i=1}^{n} X_{i-j}\,\varphi(R_{i\mathbf{u}}/(n+1)), \quad \mathbf{u}\in\mathbb{R}^p,\ 1\le j\le p,$$

(8) $$\sup_{1\le j\le p,\,\mathbf{u}} n^{1/2}|S_j(\mathbf{u})-\tilde{S}_j(\mathbf{u})| \le p \max_{1-p\le k\le 0}|X_k|\,\|\varphi\|_\infty\, n^{-1/2} \longrightarrow 0, \text{ a.s..}$$

Thus it suffices to prove the theorem with $\{S_j\}$ replaced by $\{\tilde{S}_j\}$. Let $\tilde{\mathcal{S}}' := (\tilde{S}_1, \ldots, \tilde{S}_p)$. Observe that

$$\tilde{\mathcal{S}}(\mathbf{u}) = n^{-1}\sum_{i=1}^{n} \mathbf{Y}_{i-1}\,\varphi(R_{i\mathbf{u}}/(n+1)), \qquad \mathbf{u}\in\mathbb{R}^p.$$

The proof is facilitated by centering $\tilde{\mathcal{S}}$. Accordingly, define

$$\mathbf{M}(\mathbf{u}) := n^{-1}\sum_{i=1}^{n} \mathbf{Y}_{i-1}[\varphi(R_{i\mathbf{u}}/(n+1)) - \bar{\varphi}], \qquad \mathbf{u}\in\mathbb{R}^p,$$

$$\hat{\mathbf{M}} := n^{-1}\sum_{i=1}^{n} \mathbf{Y}_{i-1}[\varphi(F(\epsilon_i)) - \bar{\varphi}].$$

As in the proof of Theorems 7.2.1, 7.2.2, let $\mathbf{M}(\mathbf{u})$, $F_n(\cdot, \mathbf{u})$, etc. stand for $\mathbf{M}(\rho+n^{-1/2}\mathbf{u})$, $F_n(\cdot, \rho+n^{-1/2}\mathbf{u})$, etc. Thus, e.g., $F_n(\cdot, \mathbf{0})$ now stands for the empirical d.f. of ϵ_i, $1\le i\le n$. Write $F_n(\cdot)$ for $F_n(\cdot, \mathbf{0})$. Recall, from the proof of Theorem 7.2.2 that $\epsilon_{i\mathbf{u}} = \epsilon_i - n^{-1/2}\mathbf{u}'\mathbf{Y}_{i-1}$, $n^{-1}R_{i\mathbf{u}} \equiv F_n(\epsilon_{i\mathbf{u}}, \mathbf{u})$. Now, let

$$e_{ni\mathbf{u}} = n^{1/2}[(R_{i\mathbf{u}}/(n+1)) - F(\epsilon_i)], \qquad 1\le i\le n,\ \mathbf{u}\in\mathbb{R}^p.$$

We first prove the

(9) **Claim:** $$\sup_{i,\mathbf{u}} n^{-1/2}|e_{ni\mathbf{u}}| = o_p(1).$$

As in the proof of Theorem 7.2.2, the supremum w.r.t. i, $\mathbf{u}$ will be over $1\le i\le n$, $\mathcal{N}(B)$, respectively, unless mentioned otherwise.

To begin with, $|[n(n+1)^{-1} - 1]| = O(n^{-1})$ implies that

(10) $$\sup_{i,\mathbf{u}} |n^{-1/2}e_{ni\mathbf{u}} - [F_n(\epsilon_{i\mathbf{u}}, \mathbf{u}) - F(\epsilon_i)]| = O(n^{-1}), \quad \text{a.s.}$$

Now, in view of (3), (4) and the discussion preceeding the stament of this theorem, it follows that $\{X_i\}$ are stationary, ergodic and hence by (c1(i)) and the Ergodic Theorem, $(1/n)\Sigma_i \mathbf{Y}_{i-1} \longrightarrow E\mathbf{Y}_0 = \mathbf{0}$. This together with (6) above, Remark 7.2.3 and (7.2.14*) imply that

$$\sup_{|x|<\infty,\ \|\mathbf{u}\|\le B} n^{1/2}|F_n(x, \mathbf{u}) - F_n(x)| = o_p(1).$$

This together with (6), (10) and (7.2.19) readily imply that

$$n^{-1/2}e_{ni\mathbf{u}} = [F_n(\epsilon_{i\mathbf{u}}) - F(\epsilon_i)] + \bar{o}_p(n^{-1/2})$$

$$(11) \qquad = [F_n(\epsilon_i) - F(\epsilon_i)] - n^{-1/2}\mathbf{u}'\mathbf{Y}_{i-1} f(\epsilon_i) + \bar{o}_p(n^{-1/2}),$$

where, $\bar{o}_p(n^{-1/2})$ is an array of processes in $(i, \mathbf{u})$ that converge to zero, uniformly in $(i, \mathbf{u})$, in probability, at a rate faster than $n^{-1/2}$.

Now the Claim (9) follows from (6), the Glivenko – Cantelli Lemma and the assumption (**F1**) that ensures $\|f\|_\infty < \infty$.

Next, define

$$\mathcal{T}(\mathbf{u}) := n^{-1} \Sigma_i \mathbf{Y}_{i-1}e_{ni\mathbf{u}}\, \dot{\varphi}(F(\epsilon_i)), \qquad \mathbf{u} \in \mathbb{R}^p.$$

Note that

$$\mathbf{M}(\mathbf{u}) = n^{-1} \Sigma_i \mathbf{Y}_{i-1} [\varphi(F(\epsilon_i) + n^{-1/2}e_{ni\mathbf{u}}) - \bar{\varphi}\,].$$

Therefore, from the uniform continuity of $\dot{\varphi}$, the facts that $n^{-1}\Sigma_i \|\mathbf{Y}_{i-1}\| = O_p(1) = n^{-1}\|\Sigma_i \mathbf{Y}_{i-1}\mathbf{Y}'_{i-1}\|$, which in turn follow from the assumption $E\epsilon^2 < \infty$ and the Erogodic Theorem, and from (9), one readily concludes that, with $U_i \equiv F(\epsilon_i)$,

$$\|n^{1/2}[\mathbf{M}(\mathbf{u}) - \hat{\mathbf{M}}] - \mathcal{T}(\mathbf{u})\|$$

$$= \|n^{-1/2}\Sigma_i \mathbf{Y}_{i-1}\{\varphi(U_i + n^{-1/2}e_{ni\mathbf{u}}) - \varphi(U_i) - n^{-1/2}e_{ni\mathbf{u}}\, \dot{\varphi}(U_i)\}\|$$

$$(12) \qquad = \bar{o}_p(1).$$

Next, we approximate $\mathcal{T}(\mathbf{u})$. Again, by the Ergodic Theorem, the independence of $\mathbf{Y}_{i-1}$ from ϵ_i, $i \ge 1$, and $E\epsilon = 0$ imply that

$$n^{-1}\Sigma_i \mathbf{Y}_{i-1}\dot{\varphi}(U_i)\, f(\epsilon_i) \longrightarrow 0, \text{ a.s.}$$

Hence by (11),

$$(13) \quad \mathcal{T}(\mathbf{u}) = n^{-1/2}\Sigma_i \mathbf{Y}_{i-1}[\{F_n(\epsilon_i) - F(\epsilon_i)\} - n^{-1/2}\mathbf{u}'\mathbf{Y}_{i-1}f(\epsilon_i)]\dot{\varphi}(U_i) + \bar{o}_p(1)$$

$$= \mathcal{V}_n - \mathbf{u}'\mathbf{L}_n + \bar{o}_p(1),$$

where now $\bar{o}_p(1)$ is a sequence of stochastic processes converging to zero,

uniformly in $\mathbf{u}$, in probability, and where

$$\mathcal{V}_n := n^{-1/2} \Sigma_i \mathbf{Y}_{i-1} [F_n(\epsilon_i) - F(\epsilon_i)] \dot{\varphi}(F(\epsilon_i)),$$
$$\mathbf{L}_n := n^{-1} \Sigma_i \mathbf{Y}_{i-1}\mathbf{Y}'_{i-1} f(\epsilon_i) \dot{\varphi}(F(\epsilon_i)).$$

Note that

$$E\mathbf{L}_n = E(n^{-1}\Sigma_i \mathbf{Y}_{i-1}\mathbf{Y}'_{i-1}) Q = \mathbf{\Sigma} Q, \qquad Q = \int f\, d\varphi(F).$$

By the Ergodic Theorem,

$$\mathbf{L}_n \longrightarrow \mathbf{\Sigma} Q, \text{ a.s..} \tag{14}$$

Our next goal is to approximate $\mathcal{V}_n$. To that effect, let $\mathcal{V}_{nj}$ denote the jth component of $\mathcal{V}_n$. Define

$$\mathcal{U}_{nj}(x) := n^{-1/2} \Sigma_i X_{i-j} \dot{\varphi}(F(\epsilon_i)) I(\epsilon_i \le x)$$
$$\nu_{nj}(x) := n^{-1/2} \Sigma_i X_{i-j} \int_{-\infty}^{x} \dot{\varphi}(F(y))\, dF(y) = n^{-1/2} \Sigma_i X_{i-j} \varphi(F(x)),$$
$$\mathcal{K}_{nj}(x) := \mathcal{U}_{nj}(x) - \nu_{nj}(x), \qquad x \in \mathbb{R},$$
$$\tilde{\mathcal{Z}}_j := \mathcal{K}_{nj}(F^{-1}(t)), \qquad 0 \le t \le 1,\ 1 \le j \le p.$$

Observe that

$$\mathcal{V}_{nj} = \int [F_n - F]\, d\,\mathcal{U}_{nj} = \int [F_n - F]\, d\mathcal{K}_{nj} + \int [F_n - F]\, d\nu_{nj}$$
$$= -\int_0^1 [\mathcal{K}_{nj}(F_n^{-1}(t)) - \mathcal{K}_{nj}(F^{-1}(t))]\, dt - \int \nu_{nj}\, d[F_n - F]$$
$$= -\int_0^1 [\tilde{\mathcal{Z}}_j(F(F_n^{-1}(t))) - \tilde{\mathcal{Z}}_j(t)]\, dt - \int \nu_{nj}\, d[F_n - F].$$

But, $\tilde{\mathcal{Z}}_j$ is a $\mathcal{Z}_j$–process of Lemma 7.3b.1 with $h = \dot{\varphi}$. Hence

$$\max_{1 \le j \le p} |\mathcal{V}_{nj} + \int \nu_{nj}\, d(F_n - F)| \tag{15}$$
$$\le \sup_{1 \le j \le p,\, 0 \le t \le 1} |\tilde{\mathcal{Z}}_j(F(F_n^{-1}(t))) - \tilde{\mathcal{Z}}_j(t)| = o_p(1),$$

by Lemma 7.3b.1 and the fact that $\sup[|F(F_n^{-1}(t)) - t|;\ 0 \le t \le 1] = o_p(1)$, which in turn follows from Lemma 3.4.1.

Next, observe that, with $\tilde{X}_j = n^{-1}\sum_{i=1}^{n} X_{i-j}$, $1 \le j \le p$,

$$\begin{aligned}\int \nu_{nj}\, d(F_n - F) &= n^{-3/2}\, \Sigma_i X_{i-j}\, \Sigma_i\, [\varphi(F(\epsilon_i)) - \overline{\varphi}\,] \\ &= \tilde{X}_j\, n^{-1/2}\, \Sigma_i\, [\varphi(F(\epsilon_i)) - \overline{\varphi}\,].\end{aligned}$$

Let $\tilde{\mathbf{X}}' := (\tilde{X}_1, \ldots, \tilde{X}_p)$ and $\hat{T} = n^{-1/2}\, \Sigma_i\, [\varphi(F(\epsilon_i)) - \overline{\varphi}\,]$. Then from (13) – (15) we obtain

(16) $$V_n = -\tilde{\mathbf{X}}\, \hat{T} + o_p(1), \qquad \mathcal{T}(\mathbf{u}) = -\tilde{\mathbf{X}}\, \hat{T} - \mathbf{\Sigma}\, \mathbf{u}\, Q + \overline{o}_p(1).$$

From (12), (16) and direct algebra one now readily concludes that

$$\begin{aligned}\mathbf{M}(\mathbf{u}) &= \hat{\mathbf{M}} - \tilde{\mathbf{X}}\, \hat{T} - \mathbf{\Sigma}\, \mathbf{u}\, Q + \overline{o}_p(1) \\ &= n^{-1/2} \sum_{i=1}^{n} (\mathbf{Y}_{i-1} - \tilde{\mathbf{X}})\, [\varphi(F(\epsilon_i)) - \overline{\varphi}\,] - \mathbf{\Sigma}\, \mathbf{u}\, Q + \overline{o}_p(1).\end{aligned}$$

Now argue as for (5) to conclude that

$$\| n^{-1/2} \sum_{i=1}^{n} (\mathbf{Y}_{i-1} - \tilde{\mathbf{X}})\, [\varphi(F(\epsilon_i)) - \overline{\varphi}\,] - \hat{\mathcal{S}} \| = o_p(1),$$

thereby completing the proof of (3). □

Remark 7.3b.1. Note that the same proof shows that under the assumed conditions, for every $0 < B < \infty$,

$$\sup_{\|\mathbf{u}\| \le B} \| \mathcal{S}(\rho + n^{-1/2}\mathbf{u}) - \mathcal{S}(\rho) + \mathbf{u}'\, \mathbf{\Sigma} \int f\, d\varphi(F) \,\| = o_p(1). \qquad \square$$

Remark 7.3b.2. Argue either as in Section 3.4 or as in Jaeckal (1972) to conclude that $\|n^{1/2}(\hat{\rho}_R - \rho)\| = O_p(1)$ and that $\|n^{1/2}(\hat{\rho}_R - \hat{\rho}_J)\| = o_p(1)$. Consequently by Theorem 7.3b.1,

(17) $$n^{1/2}(\hat{\rho}_R - \rho) = n^{1/2}(\tilde{\rho}_J - \rho) + o_p(1) = Q^{-1}\, \mathbf{\Sigma}^{-1}\, \hat{\mathcal{S}} + o_p(1).$$

Observe that $\hat{\mathcal{S}}$ is a vector of square integrable mean zero martingales with $E\hat{\mathcal{S}}\hat{\mathcal{S}}' = \sigma_\varphi^2\, \Sigma$, $\sigma_\varphi^2 := \mathrm{Var.}(\varphi(U))$. Thus, by the routine Cramer–Wold device and by Lemma A.3 in the Appendix, one readily obtains

(18) $$\hat{\mathcal{S}} \xrightarrow[d]{} N(0, \sigma_\varphi^2\, \mathbf{\Sigma}),$$

(19) $n^{1/2}(\hat{\rho}_R - \rho) \xrightarrow[d]{} N(0, \mathcal{V}),\ n^{1/2}(\tilde{\rho}_J - \rho) \xrightarrow[d]{} N(0, \mathcal{V}),\ \mathcal{V} = Q^{-2}\sigma_\varphi^2\, \Sigma^{-1}$. □

Remark 7.3b.3. See the recent paper of Koul and Ossiander (1992) for an extension of the above results to any $\varphi \in \mathscr{C}$ of (3.2.1). □

7.3c. ESTIMATION OF $Q(f) := \int f\, d\varphi(F)$.

As is evident from (7.3b.19), the rank analysis of an AR(p) model via the above R-estimators will need a consistent estimator of the functional Q. In this subsection we give two classes of consistent estimators of this functional in the AR(p) model (7.3b.3), (7.3b.4). One class of estimators is obtained by replacing f and F in Q by a kernel density estimator and the empirical d.f. based on the estimated residuals, respectively. This is analogous to the class of estimators discussed in Theorem 4.5.3. The other class is an analogue of the class of estimators discussed in Theorem 4.5.1 in connection with the linear regression setup.

Accordingly, let $\tilde{\rho}$ be an estimator of ρ, K be a probability density in $\mathbb{R}$, h_n be a sequence of positive numbers, $h_n \longrightarrow 0$ and define, for $x \in \mathbb{R}$,

$$\tilde{\epsilon}_i := X_i - \tilde{\rho}' Y_{i-1},\ 1 \le i \le n; \qquad \tilde{F}_n(x) := F_n(x, \tilde{\rho}) = n^{-1} \Sigma_i\, I(\tilde{\epsilon}_i \le x),$$

$$\tilde{f}_n(x) := (nh_n)^{-1} \Sigma_i\, K\Big(\frac{x - \tilde{\epsilon}_i}{h_n}\Big), \qquad f_n(x) := (nh_n)^{-1} \Sigma_i\, K\Big(\frac{x - \epsilon_i}{h_n}\Big).$$

Finally, let

$$\tilde{\mathcal{Q}}_n := \int \tilde{f}_n\, d\varphi(\tilde{F}_n).$$

Theorem 7.3c.1. *In addition to* (7.3.b3), (7.3.b4), *assume that* $E\epsilon_1 = 0$, $E\epsilon_1^2 < \infty$. *Moreover, assume that* **(F1)**, **(F2)** *and the following conditions hold.*

(i) $\varphi \in \mathscr{C} := \{\varphi\colon \varphi$ *a nondecreasing function on* $[0, 1]$, $\varphi(0) = 0,\ \varphi(1) = 1\}$.

(ii) $h_n > 0;\ h_n \longrightarrow 0,\ n^{1/2} h_n \longrightarrow \infty$.

(iii) K *is absolutely continuous with its a.e derivative* $\dot{K}$ *satisfying* $\int |\dot{K}| < \infty$.

(iv) $\|n^{1/2}(\tilde{\rho} - \rho)\| = O_p(1)$.

Then,

(1) $$\sup_{\varphi \in \mathscr{C}} |\tilde{\mathcal{Q}}_n - Q(f)| = o_p(1).$$

Proof. The proof is similar to that of Theorem 4.5.3, so we shall be brief, indicating only one major difference. Unlike in the linear regression setup, i.e., unlike (4.5.11), here we have from Remark 7.2.3,

$$\sup_x n^{1/2}|\tilde{F}_n(x) - F_n(x)| = o_p(1), \quad \text{where } F_n(x) \equiv F_n(x, \rho). \tag{2}$$

In other words the linearity term involving $n^{1/2}(\tilde{\rho} - \rho)$ is not present in the approximation of $\tilde{F}_n$. Proceeding as in the proof of Theorem 4.5.3, (2) will yield

$$\begin{aligned} \|\tilde{f}_n - f_n\|_\infty &\le (n^{1/2}h_n)^{-1} \cdot \|n^{1/2}[\tilde{F}_n - F_n]\|_\infty \cdot \int |\dot{K}| \\ &= o_p((n^{1/2}h_n)^{-1}) = o_p(1). \end{aligned}$$

Compare this with (4.5.19) where $O_p((n^{1/2}h_n)^{-1})$ appears instead of $o_p((n^{1/2}h_n)^{-1})$. Rest of the proof is exactly the same as there with the proviso that one uses (2) instead of (4.5.11), whenever needed. □

The reader may wish to modify the above proof to see that $\tilde{Q}_n$ continues to be consistent for Q even when $E\epsilon \neq 0$, so that the term that is linear in $n^{1/2}(\tilde{\rho} - \rho)$ is now present in the expansion of $\tilde{F}_n$.

We shall now describe an analogue of $\hat{Q}_n^\alpha$ of (4.5.6). The motivation is the same as in Section 4.5, so we shall be brief on that also. Accordingly, let

$$\tilde{p}(y) := \int [\tilde{F}_n(y+x) - \tilde{F}_n(-y+x)]\, d\, \varphi(\tilde{F}_n(x)), \qquad y \ge 0.$$

Observe that $\tilde{p}$ is an estimator of the d.f. of the absolute difference $|\epsilon - \eta|$, where ϵ and η are independent r.v.'s with respective d.f.'s F and $\varphi(F)$. As in Section 4.5, one can use the following representation for the computational purposes.

$$\tilde{p}(y) = n^{-1} \sum_{j=1}^{n} [\varphi(j/n) - \varphi((j-1)/n)] \sum_{i=1}^{n} I(|\tilde{\epsilon}_{(i)} - \tilde{\epsilon}_{(j)}| \le y), \qquad y \ge 0,$$

where $\{\tilde{\epsilon}_{(i)}\}$ are the ordered residuals $\{\tilde{\epsilon}_i\}$ from the smallest to the largest.

Now let $\tilde{t}_n^\alpha$ denote an αth percentile of the d.f. $\tilde{p}(y)$ and define

$$\tilde{Q}_n^\alpha = n^{1/2}\, \tilde{p}(n^{-1/2}\, \tilde{t}_n^\alpha) \,/\, 2\tilde{t}_n^\alpha, \qquad 0 < \alpha < 1.$$

The consistency of these estimators may be proved using the method of the proof of Theorem 4.5.1 and the results given in Corollary 7.2.1. The discussion about the choice of α etc. that appears in Remark 4.5.1 is also pertinent here.

Another class of estimators is obtained by modifying $\tilde{Q}_n$ by replacing $\hat{F}_n$ by the estimator $\overline{F}_n(x) = \int \tilde{f}_n(y)\, I(-\infty<y\le x)\, dy$. The consistency of these estimators can be also proved by the help of Corollary 7.2.1. □

7.3d. PROOF OF LEMMA 7.3b.3.

The proof of Lemma 7.3b.3 is similar to that of Theorem 2.2a.1(i) and will be a consequence of the following *two* lemmas.

Lemma 7.3d.1. *In addition to* (7.3b.3), (7.3b.4) *and* (7.3b.c1) *assume that the following hold:*

(d1) *The d.f.* F *is continuous and strictly increasing.*

(d2) *The function h on* $[0, 1]$ *to* $\mathbb{R}$ *is nonnegative and* $\int |h(t)|^4 dt < \infty$.

Then the following hold:

(A) *For any* $0 \le u < v < w \le 1$ *and for all* $1 \le j \le p$,

(1) $\limsup_n E\{ \mathcal{Z}_j(v) - \mathcal{Z}_j(u)\}^2\{(\mathcal{Z}_j(w) - \mathcal{Z}_j(v)\}^2 \le C\, m_1 m_2$,

where $m_1 := \int_u^v h^2(t)\, dt$, $m_2 := \int_v^w h^2(t)\, dt$, C *is a constant given in* (19) *below.*

(B) *For any* $0 \le u < v \le 1$, *and for* $1 \le j \le p$,

$$\limsup_n E\,\{ \mathcal{Z}_j(v) - \mathcal{Z}_j(u)\}^4 \le C\, m_1^2.$$

Proof. (A). Since u, v, w, are fixed, we shall suppress these entities in the notation. Let $\mathcal{F}_k := \sigma$–field$\{\epsilon_i;\ i \le k\}$, $k = 0, \pm1, \pm2, \ldots$. Further, to simplify writing let $x = F^{-1}(u)$, $y = F^{-1}(v)$ and $z = F^{-1}(w)$ and define

(2) $p_1 = H(v) - H(u)$, $p_2 = H(w) - H(v)$; $q_j = 1 - p_j$, $j = 1, 2$,

$$\alpha_i := h(F(\epsilon_i))I(x < \epsilon_i \le y) - p_1, \quad \beta_i := h(F(\epsilon_i))I(y < \epsilon_i \le z) - p_2.$$

Then

(3) $\{ \mathcal{Z}_j(v) - \mathcal{Z}_j(u)\}^2\{ \mathcal{Z}_j(w) - \mathcal{Z}_j(v)\}^2 = n^{-2}(\Sigma_i X_{i-j}\alpha_i)^2 \cdot (\Sigma_r X_{r-j}\beta_r)^2$.

In carrying out the computations that follow we have repeatedly used the following facts: α_i, β_i are centered; $\alpha_i\beta_i$ are $\mathcal{F}_{k-1}$ measurable for all $i < k$ and X_{i-j} is $\mathcal{F}_{i-1}$ measurable and independent of ϵ_i, $i \ge 1$. Thus,

(4) $\quad E\,\alpha_i = 0 = E\,\beta_k,$ for all i,k.

$$EX_{i-j}X_{k-j}\alpha_i\beta_k = E[X_{i-j}X_{k-j}\alpha_i E(\beta_k|\mathcal{F}_{k-1})] = E[X_{i-j}X_{k-j}\alpha_i]\,E(\beta_k) = 0, \quad i<k;$$

$$EX_{i-j}X_{k-j}X^2_{r-j}\alpha_i\,\alpha_k\,\beta_r^2 = E[X_{i-j}X_{k-j}X^2_{r-j}\beta_r^2\,\alpha_i E(\alpha_k|\mathcal{F}_{k-1})] = 0, \qquad i,r<k.$$

Using facts like these one can write

(5) $\quad E(\Sigma_i\, X_{i-j}\,\alpha_i)^2(\Sigma_r\, X_{r-j}\beta_r)^2$

$$\begin{aligned} &= \sum_i E\,X^4_{i-j}\,\alpha_i^2\,\beta_i^2 + \sum_{i\neq r}\sum EX^2_{i-j}\,X^2_{r-j}\alpha_i^2\,\beta_r^2 \\ &\quad + 4\sum_{i<k}\sum E\,X^2_{i-j}X^2_{k-j}\,\alpha_i\,\beta_i\,\alpha_k\,\beta_k \\ &\quad + 2\sum_{i<k<r}\sum\sum E\,X_{i-j}\,X_{k-j}\,X^2_{r-j}(\alpha_i\,\alpha_k\,\beta_r^2 + \beta_i\,\beta_k\,\alpha_r^2) \\ &\quad + 4\sum_{i<r<k}\sum\sum E\,X_{i-j}\,X_{r-j}\,X^2_{k-j}\,\alpha_i\,\beta_r\,\alpha_k\,\beta_k \\ &\quad + 4\sum_{r<i<k}\sum\sum E\,X_{i-j}\,X_{r-j}\,X^2_{k-j}\,\alpha_i\,\beta_r\,\alpha_k\,\beta_k \\ &= T_1 + T_2 + 4\,T_3 + 2(T_4 + T_5) + 4(T_6 + T_7), \text{ say.}\end{aligned}$$

We shall now show that $n^{-2}T_j \longrightarrow 0$, for $j = 1, 4, 5, 6, 7$, and that $\limsup n^{-2}(T_2 + 4\,T_3) \leq C\,m_1m_2$. The basic idea of the proof is to exploit the hierarchal nature of the process. Observe that had the underlying observations been independent then T_j would have been equal to zero for $j = 4, 5, 6, 7$. However, under (7.3b.3), $\{X_j\}$ are not independent but asymptotically behave like independent r.v.'s. This is the reason to expect $n^{-2}T_j \longrightarrow 0$ for $j = 4, 5, 6, 7$.

The details of the proof of $n^{-2}T_j$ tending to zero for each $j = 4, 5, 6, 7$ are elementary and cumbersome but similar. So the details will be given only for $n^{-2}T_7 \longrightarrow 0$. To this effect, observe that

$$E(X_{i-j}X_{r-j}X^2_{k-j}\alpha_i\beta_r\alpha_k\beta_k) = E\,X_{i-j}X_{r-j}X^2_{k-j}\alpha_i\beta_r E(\alpha_k\,\beta_k|\mathcal{F}_{k-1}), \quad i,r<k.$$

Moreover, $\{\epsilon_i\}$ i. i. d. implies that for all $k \geq 1$,

$$E(\alpha_k\beta_k|\mathcal{F}_{k-1}) = (1-p_1)p_1(-p_2)+(-p_1)(1-p_2)p_2+p_1p_2(1-p_1-p_2) = -p_1p_2,$$

and, in addition, $E\epsilon = 0$ implies that $EX_{i-j}X_{r-j}X^2_{k-j}\alpha_i \equiv 0$, $r < i$, $k-j \leq i-1$. Therefore,

(6) $$n^{-2}T_7 = -n^{-2}p_1p_2\{\sum_{r<i}\sum EX_{i-j}\, X_{r-j}\, X_i^2\, \alpha_i\, \beta_r + \sum_{r<i}\sum \sum_{k-j>i} EX_{i-j}\, X_{r-j}\, X_{k-j}^2\, \alpha_i\, \beta_r\}$$
$$= -\, p_1p_2\, n^{-2}\{T_{71} + T_{72}\}, \qquad \text{say.}$$

Now, for a convenient reference we rewrite (7.3b.5) as

(7) $$X_i = \sum_{k\le i} \theta_{i-k}\, \epsilon_k, \qquad i \ge 1, \text{ a.s.,}$$

where, as in (7.3b.5), the unspecified lower limit on the index of summation is $-\infty$, and $\{\theta_k\}$ are real numbers satisfying $\theta_0 = 1$, $\Delta_1 < \infty$, with

$$\Delta_q := \Sigma_{k\ge 0}\, |\theta_k|^q, \quad q \ge 1.$$

Note that $\sup_k |\theta_k| \le \Delta_1$ and hence $\Delta_1 < \infty$ implies that

(8) $$\Delta_r \le \Delta_1^q < \infty, \qquad \text{for all } q \ge 1.$$

Next, define

(9) $$A_{m,n} := \sum_{r\le n} \theta_{m-r}\, \epsilon_r,$$

$$H_{m,k}^n := \sum_{r=k}^{n} \theta_{m-r}\, \epsilon_r, \qquad 0 \le n \le m < \infty,\ k \le n.$$

$$a_r := E(\alpha\, \epsilon^r), \quad b_r := E(\beta\, \epsilon^r), \quad \mu_r := E\epsilon^r, \qquad 1 \le r \le 4,$$

$$\sigma_1^2 := \mathrm{Var.}(\alpha), \quad \sigma_2^2 := \mathrm{Var.}(\beta)$$

where α, β are copies of α_1, β_1. Observe that

(10) $$H_{m,k}^n = A_{m,n} - A_{m,k-1}, \qquad k \le n \le m,$$

$$X_i = A_{i,i} = A_{i,i-j} + H_{i,i-j+1}^{i-1} + \epsilon_i = A_{i,i-1} + \epsilon_i, \qquad \forall\ i,$$

$$\sigma_k^2 \le m_k, \quad \text{where } m_k \text{ is as in (1)}, \qquad k = 1, 2.$$

Morover, $\{\epsilon_i\}$ i.i.d., $E(\epsilon) = 0$, and (8) imply that for all $n \le m < \infty$,

(11) $$EA_{m,n}^2 = \sum_{r\ge m-n} \theta_r^2\, \mu_2 \le \mu_2\, \Delta_1^2 < \infty,$$

$$EA_{m,n}^4 = \sum_{k\ge m-n} \theta_k^4 \mu_4 + 3\sum_{k\ge m-n}\ \sum_{r\ge m-n, r\ne k} \theta_k^2\theta_r^2\mu_2 \le (\mu_4+\mu_2)\Delta_1^4 < \infty.$$

For the same reasons, from (10) it follows that

$$E\{X_i^2\ \alpha_i | \mathcal{F}_{i-1}\} = 2\ A_{i,i-1}\, a_1 + a_2, \qquad \text{for all } i.$$

Use this and argue as for (6) to obtain

$$\begin{aligned} T_{71} = {} & \Sigma_r\, E\, X_{r-j}\beta_r\, \{2a_1\, L_r + a_2\, X_r\} \\ & + \sum_{r<i;\, i-j\geq r+1} \sum E\, X_{r-j}\beta_r\, \{2a_1 L_i + a_2 X_{i-j}\} \\ = {} & T_{711} + T_{712}, \quad \text{say, where} \quad L_i := X_{i-j}\, A_{i,i-1}. \end{aligned}$$

The C–S inequality, the staionarity of the process $\{X_i\}$ and (11) imply that for all r, j,

$$E|X_{r-j}\beta_r\, L_{r+j}| \leq \{E(X_{r-j}\beta_r)^2\, EL_{r+j}^2\}^{1/2} \leq D_4\, C_4 < \infty,$$

$$E|X_{r-j}\beta_r\, L_r| \leq D_4\, C_4 < \infty,$$

where D_k, C_k are constants depending on the kth moment of $h(U)$ and the kth moment of ϵ and Δ_k, respectively, $1 \leq k \leq 4$. These facts imply that

$$n^{-2}|T_{711}| = O(n^{-1}) = o(1). \tag{12}$$

Next, **to handle** T_{712}, use (11) to obtain that for $i-j \geq r+1$,

$$L_i = \{A_{i-j,r-1} + \theta_{i-j-r}\, \epsilon_r + H_{i-j,r+1}^{i-j}\}\{A_{i,r-1} + \theta_{i-r}\, \epsilon_r + H_{i,r+1}^{i-1}\}.$$

Use the above type of conditioning argument to obtain that

$$EX_{r-j}\, \beta_r\, L_i = EX_{r-j}\, \{[\theta_{i-r}\, A_{i-j,r-1} + \theta_{i-j-r}\, A_{i,r-1}]\, b_1 + \theta_{i-j-r}\, \theta_{i-r}\, b_2\},$$

$$EX_{r-j}\, \beta_r\, X_{i-j} = EX_{r-j}\{\theta_{i-j-r}^2\, b_2 + 2\theta_{i-j-r}\, A_{i-j,r-1}\, b_1\}, \qquad i-j \geq r+1.$$

Use these facts together with (11) and an argument like the one that led to (12) to conclude that $n^{-2}|T_{712}| = O(n^{-2})$. This and (12) yield

$$n^{-2}|T_{71}| = O(n^{-1}) = o(1). \tag{13}$$

Now we **turn to** $\mathbf{T_{72}}$. Using (10) write

$$X_{k-j} = A_{k-j,i-1} + \theta_{k-j-i}\, \epsilon_i + H_{k-j,i+1}^{k-j}, \qquad k-j \geq i+1,$$

and use arguments like those above to obtain that

$$E\{X_{k-j}^2\ \alpha_i | \mathcal{F}_{i-1}\} = \theta_{k-j-i}^2\, a_2 + 2A_{k-j,i-1}\, \theta_{k-j-i}\, a_1, \qquad k-j \geq i+1,$$

so that

(14) $$T_{72} = \sum_{r<i<k;\,k-j\geq i+1}\sum\sum EX_{r-j}\,\beta_r\,X_{i-j}\,\{\theta^2_{k-j-i}\,a_2 + 2A_{k-j,i-1}\,\theta_{k-j-i}\,a_1\}$$
$$= a_2\,T_{721} + 2a_1\,T_{722}, \qquad \text{say.}$$

Arguing as above and using the stationarity and the fact that $EX_0 = 0$, one obtains

$$EX_{r-j}\,\beta_r\,X_{i-j} = EX_{r-j}\,\beta_r\,\{A_{i-j,r-1} + \theta_{i-j-r}\,\epsilon_r + H^{i-j}_{i-j,r+1}\}$$
$$= EX_{r-j}\,\theta_{i-j-r}\,b_1 = 0, \qquad i-j \geq r+1.$$

Thus,

$$|T_{721}| \leq a_2\,d_2 \sum_{r<i}\sum |EX_{r-j}\,\beta_r\,X_{i-j}| = 0.$$

Similar arguments show that $|n^{-2}T_{722}| = o(1)$ thereby completing the proof of $n^{-2}|T_{72}| = o(1)$. This together with (13) shows that

$$n^{-2}|T_7| = o(1).$$

Now **consider** T_2: Rewrite

$$T_2 = (\sum_{i<r}\sum + \sum_{r<i}\sum)\,(EX^2_{i-j}\,X^2_{r-j}\,\alpha^2_i\,\beta^2_r) = T_{21} + T_{22}, \quad \text{say.}$$

Again, by a conditionning argument,

(15) $$T_{21} = \sigma^2_2 \cdot \sum_{i<r}\sum EX^2_{i-j}\,X^2_{r-j}\,\alpha^2_i$$
$$= \sigma^2_2 \cdot (\sum_{i<r;\,r-i\leq j}\sum + \sum_{i<r;\,r-i\geq j+1}\sum)(EX^2_{i-j}\,X^2_{r-j}\,\alpha^2_i)$$
$$= \sigma^2_2 \cdot \{T_{211} + T_{212}\}, \qquad \text{say.}$$

Again, the C–S inequality, the stationarity of the process $\{X_i\}$, the assumptions (7.3b.c1) and (d2) imply that $0 < T_{211} \leq j\cdot n\cdot EX^4_0 = O(n)$, by (8) and (10), so that

(16) $$n^{-2}T_{211} = O(n^{-1}) = o(1).$$

Next, argue as for (14) to obtain

$$T_{212} = \sum_{i<r;\,r-i\geq j+1}\sum EX^2_{i-j}\,X^2_{r-j}\,\alpha^2_i$$
$$= \sum_{i<r;\,r-i\geq j+1}\sum EX^2_{i-j}\,\alpha^2_i\,\{A_{r-j,i-1} + \theta_{r-j-i}\,\epsilon_i + H^{r-j}_{r-j,i+1}\}^2$$

$$= \sigma_1^2 \cdot \sum_{i<r;\, r-i\geq j+1}\sum EX_{i-j}^2 \{A_{r-j,i-1}^2 + \sum_{m=i+1}^{r-j} \theta_{r-j-m}^2 E\epsilon^2\}$$

$$+ 2c \sum_{i<r;\, r-i\geq j+1}\sum EX_{i-j}^2 A_{r-j,i-1} \theta_{r-j-i}$$

(17) $$= \sigma_1^2 \cdot B_1 + 2c \cdot B_2 + O(n), \quad \text{say}, \qquad c = E(\epsilon\alpha)^2.$$

The C–S inequality and (11) yield that

(18) $$n^{-2} B_1 \leq C < \infty, \text{ for all } n \geq 1,$$

where C is a constant depending on μ_4 and Δ_1. A similar argument shows that $|B_2| = O(n^{-1})$. This together with (17) and (18) yield that

(19) $$\limsup_n n^{-2} |T_{212}| \leq C\, \sigma_1^2.$$

Hence, from (17) – (19) one readily obtains

$$\limsup_n n^{-2} |T_{21}| \leq C\, \sigma_1^2\, \sigma_2^2 \leq C\, m_1 m_2, \qquad \text{by (10)}.$$

Similarly, one concludes a similar result for T_{22} thereby enabling one to conclude

(20) $$\limsup_n n^{-2} |T_2| \leq C\, m_1 m_2, \text{ where } C \text{ is as in (18)}.$$

Finally, consider $n^{-2} T_3$: By arguments similar to those above we obtain

(21) $$n^{-2} T_3 = -n^{-2} \sum_{i<r}\sum EX_{i-j}^2\, X_{r-j}^2\, \alpha_i\, \beta_i\, p_1 p_2$$

$$= -p_1 p_2\, n^{-2} \sum_{i<r;\, r-i\geq j+1}\sum EX_{i-j}^2\, \alpha_i\, \beta_i\, X_{r-j}^2 + O(n^{-1}).$$

Let $c_r = E(\alpha\beta\, \epsilon^r)$. Use (10) and procced as before to obtain

$$EX_{i-j}^2\, \alpha_i\, \beta_i\, X_{r-j}^2 = EX_{i-j}^2\{-p_1 p_2 \,(A_{r-j,i-1}^2 + \mu_2 \sum_{m=i+1}^{r-j} \theta_{r-j-m}^2)$$

$$+ \theta_{r-j-i}^2\, c_2 + 2c_1\, A_{r-j,i-1}\, \theta_{r-j-i}\}.$$

Combine this with (21), argue as above using (11) and the C–S inequality, to obtain

$$n^{-2} T_3 = (p_1 p_2)^2 n^{-2} \sum_{i<r;\, r-i\geq j+1}\sum E\, X_{i-j}^2 [A_{r-j,i-1}^2 + \mu_2 \sum_{m=i+1}^{r-j} \theta_{r-j-m}^2] + O(n^{-1}).$$

Another application of (11) yields

$$\limsup_n |n^{-2}T_3| \le C\,(p_1p_2)^2 \le C\, m_1\, m_2\,,$$

where C is as in (18) above, because

$$p_1^2 = \Big\{\int_u^v h(t)\,dt\Big\}^2 \le (v-u)\int_u^v h^2(t)\,dt \le m_1$$

$$p_2^2 = \Big\{\int_v^w h(t)\,dt\Big\}^2 \le (w-v)\int_v^w h^2(t)\,dt \le m_2.$$

The proof of (A) is now terminated.

PROOF of (B). Fix j and define, for $r \ge 1$, $k \ge 1$,

$$u_{rk} = E\{(X_{k-j}\,\alpha_k)^r\,|\,\mathcal{F}_{k-1}\} = X_{k-j}^r\, E(\alpha_k^r\,|\,\mathcal{F}_{k-1});$$

$$U_{rk} = \sum_{i=1}^{k} u_{ri}; \qquad\qquad S_k = \sum_{i=1}^{k} X_{i-j}\,\alpha_i.$$

Now observe that

$$\mathcal{Z}_j(v) - \mathcal{Z}_j(u) = n^{-2}S_n. \tag{22}$$

Because $\{X_{i-j}\,\alpha_i\}$ are conditionally centered, gives $\mathcal{F}_{i-1}$, it readily follows that $\{S_n, \mathcal{F}_n\}$ is a mean zero martingale. Therefore, from Chow, Robbin and Teicher (1964),

$$E\,S_n^4 = E\{U_{4n} + 4\,S_n\,U_{3n} + 6\,S_n^2\,U_{2n} - 6\sum_{j=1}^{n} u_{2j}\,U_{2j}\}. \tag{23}$$

But,

$$E\sum_{k=1}^{n} u_{2k}\,U_{2k} = E\,\{\sum_{k=1}^{n} E(X_{k-j}^2\,\alpha_k^2\,|\,\mathcal{F}_{k-1})\cdot\sum_{i=1}^{k} E(X_{i-j}^2\,\alpha_i^2\,|\,\mathcal{F}_{i-1})\}$$
$$= E\sum_{k=1}^{n} X_{k-j}^2\,\sigma_1^2\cdot\sum_{i=1}^{k} X_{i-j}^2\,\sigma_1^2 = \sum\sum_{i\le k} E\,X_{i-j}^2\,X_{k-j}^2\cdot\sigma_1^4,$$

$$E\,S_nU_{3n} = E\{\sum_{i=1}^{n} X_{i-j}\,\alpha_i\cdot\sum_{k=1}^{n} E((X_{k-j}\,\alpha_k)^3\,|\,\mathcal{F}_{k-1})\}$$
$$= \sum\sum_{i\le k-j} EX_{i-j}\,X_{k-j}^3\,\alpha_i\cdot E(\alpha^3),$$

$$E\,S_n^2\,U_{2n} = \sigma_1^2\cdot E\{(\sum_i X_{i-j}^2\,\alpha_i^2 + 2\sum\sum_{i<r} X_{i-j}X_{r-j}\alpha_i\alpha_r)(\sum_k X_{k-j}^2)\}$$
$$= \sigma_1^2\cdot[\sum\sum_{i\le k-j} EX_{i-j}^2\,\alpha_i^2\,X_{k-j}^2 + \sigma_1^2\cdot\sum\sum_{k-j\le i-1} E\,X_{i-j}^2\,X_{k-j}^2$$
$$+ 2\sum\sum\sum_{i<r<k} EX_{i-j}\,X_{r-j}\,X_{k-j}^2\,\alpha_i\,\alpha_r],$$

$$\mathrm{E}\,U_{4n} = \sum_i \mathrm{E}\,X_{i-j}^4 \cdot \mathrm{E}(\alpha^4).$$

Combine the above with (23) to obtain

$$\begin{aligned} n^{-2}\mathrm{E}\,S_n^4 = n^{-2}\{ & \sum_{i=1}^n \mathrm{E}X_{i-j}^4 \cdot \mathrm{E}(\alpha^4) + 4\sum_{i\le k-j}\sum \mathrm{E}X_{i-j}X_{k-j}^3\,\alpha_i \cdot \mathrm{E}(\alpha^3) \\ & + 6\sigma_1^2 \cdot [\sum_{i\le k-j}\sum \mathrm{E}X_{i-j}^2\alpha_i^2X_{k-j}^2 + \sigma_1^2 \cdot \sum_{i<k\le j+i-1}\sum \mathrm{E}X_{i-j}^2X_{k-j}^2 \\ & + 2\sum_{i<r<k}\sum\sum \mathrm{E}X_{i-j}\,X_{r-j}\,X_{k-j}^2\,\alpha_i\,\alpha_r]\}. \\ = n^{-2}\{ & K_1 + 4\,K_2 + 6\,\sigma_1^2\,[K_3 + \sigma_1^2\,K_4 + 2\,K_5]\}, \qquad \text{say.} \end{aligned}$$

Arguing as for the proof of (A), one can show that $n^{-2}K_j \longrightarrow 0$, $j = 1, 2, 5$, that $\limsup n^{-2}|K_3| \le C\,\sigma_1^2$, and that $\limsup n^{-2}K_4 \le C$. Hence (B). □

Lemma 7.3d.2. *In addition to* (7.3b.3) *and* (7.3b.4) *assume that* $\mathrm{E}\epsilon = 0$, $\mathrm{E}\epsilon^2 < \infty$, *and* $\mathrm{E}h^2(U) < \infty$. *Then the finite dimensional distribution of* $\mathcal{Z}_j$, *for every* $1 \le j \le p$, *converges weakly to that of* $\{\mathrm{E}(X_0)^2\}^{1/2}B(\cdot)$, *where* B *is the Brownian motion in* $\mathbb{C}[0, 1]$ *with the covariance function* $H(u) - H(u)H(v)$, $0 \le u \le v \le 1$.

Proof. The proof uses Corollary 3.1 of Hall and Heyde (1980; p 58)(see Lemma A.3 in the Appendix) and the Cramer–Wold device. Accordingly, fix j and let $0 \le u_1 \le u_2 \le \dots \le u_r \le 1$, $\theta \in \mathbb{R}^r$. Define

$$\alpha_n(\epsilon_i) := \sum_{k=1}^r \lambda_k\,\{h(F(\epsilon_i))I(F(\epsilon_i) \le u_k) - G(u_k)\},$$

$$\xi_{ni} := n^{-1/2}\,X_{i-j}\,\alpha_n(\epsilon_i), \qquad \mathcal{S}_{ni} := \sum_{k=1}^i \xi_{nk}, \qquad 1 \le i \le n.$$

Note that $\mathcal{S}_{nn} = \sum_{k=1}^r \theta_k\,\mathcal{Z}_j(u_k)$. Because of the given assumptions, and because ξ_{ni} is conditionally centered, given $\mathcal{F}_{i-1}$, $\{(\mathcal{S}_{ni}, \mathcal{F}_{i-1}), 1 \le i \le n\}$ is a mean zero square integrable martingale array. Next, for a $\theta > 0$, by the C-S inequality,

$$\begin{aligned} & \sum_{i=1}^n \mathrm{E}[\xi_{ni}^2\,I(|\xi_{ni}| > \theta)|\mathcal{F}_{i-1}] \\ & \quad = n^{-1}\sum_{i=1}^n \mathrm{E}X_{i-j}^2\,\{\mathrm{E}[\alpha_n^2(\epsilon_i)I(|X_{i-j}\,\alpha_n(\epsilon_i)| > \theta n^{1/2})|\mathcal{F}_{i-1}]\} \\ & \quad \le n^{-1}\sum_{i=1}^n \mathrm{E}X_{i-j}^2\,P^{1/2}(|X_{i-j}\,\alpha_n(\epsilon_i)| > \theta n^{1/2}|\mathcal{F}_{i-1}) \cdot D_{4,r} \end{aligned}$$

$$\le (\theta n^{3/2})^{-1} \sum_{i=1}^{n} E|X_{i-j}|^3 E^{1/2}[|\alpha_n(\epsilon_i)|\mathcal{F}_{i-1}]\cdot D_{4,r}$$

$$\le C_3 D_{4,r}(\theta n^{1/2})^{-1} = o(1)$$

where, in the above, $D_{4,r}$ is a constant dependin on r, D_4 and θ and C_3 is a constant dependin on μ_3 and d_1.

Next, from the definition of H in terms of h one readily sees that

$$\begin{aligned}\sum_{i=1}^{n} E(\xi_{ni}^2|\mathcal{F}_{i-1}) &= n^{-1} \sum_{i=1}^{n} X_{i-j}^2\, E[\alpha_n^2(\epsilon_i)|\mathcal{F}_{i-1}] \\ &= n^{-1} \sum_{i=1}^{n} X_{i-j}^2 \sum_{k=1}^{r} \sum_{m=1}^{r} \lambda_k\lambda_m\, [G(u_k\wedge u_m) - G(u_k)\, G(u_m)] \\ &= E(X_0^2) \sum_{k=1}^{r} \sum_{m=1}^{r} \lambda_k\lambda_m[G(u_k\wedge u_m) - G(u_k)G(u_m)] + o_p(1),\end{aligned}$$

by the Ergodic Theorem.

The above calculations show that $\{\mathcal{S}_{ni}, \mathcal{F}_{i-1}, 1 \le i \le n\}$ satisfy the conditions of Lemma A.3 and hence $\mathcal{S}_{nn}$ converges weakly to an appropriate normal r.v. This completes the proof of the Lemma. □

Proof of Lemma 7.3b.3. In view of the Lemmas 7.3d.1(A) and 7.3d.2 above, the proof uses Lemmas A.1, A.2 and Theorem A.1 in the Appendix and is exactly like that of Theorem 2.2a.1(i). □

7.4. M.D. ESTIMATION

In this section we shall discuss two classes of m.d. estimators. They are the analogues of the classes of estimators defined in the linear regression setup at (5.2.11) and (5.2.20). To be precise, consider the autoregression model (7.1.1) and define, for a $G \in \mathbb{D}\mathcal{I}(\mathbb{R})$,

$$\text{(1)}\quad K_g(\mathbf{t}) = \sum_{j=1}^{p} \int [n^{-1/2} \sum_{i=1}^{n} g(X_{i-j})\{I(X_i \le x + \mathbf{t}'\mathbf{Y}_{i-1}) - F(y)\}]^2\, dG(x),$$

$$K_g^+(\mathbf{t}) = \sum_{j=1}^{p} \int [n^{-1/2} \sum_{i=1}^{n} g(X_{i-j})\{I(X_i \le x + \mathbf{t}'\mathbf{Y}_{i-1}) - I(-X_i < x - \mathbf{t}'\mathbf{Y}_{i-1})\}]^2\, dG(x), \quad \mathbf{t} \in \mathbb{R}^p,$$

In the case the error d.f. F is *known*, define a class of m.d. estimators of ρ to be

$$\text{(3)}\qquad \hat{\rho}_g := \operatorname{argmin}\{K_g(\mathbf{t});\ \mathbf{t} \in \mathbb{R}^p\}.$$

In the case the error distribution is *unknown but symmetric around* 0, define a class of m.d. estimators of ρ to be

$$\rho_g^+ := \text{argmin}\{K_g^+(\mathbf{t});\ \mathbf{t} \in \mathbb{R}^p\}. \tag{4}$$

Note that the role played by the vectors $\{n^{-1/2}[g(X_{i-1}), g(X_{i-2}), \ldots, g(X_{i-p})];\ 1 \le i \le n\}$ is similar to that of the vectors $\{\mathbf{d}_{ni};\ 1 \le i \le n\}$ of Chapter 5. To put it in matrices, the precise analogue of $\mathbf{D}$ is the matrix $n^{-1/2}\mathcal{G}$, where $\mathcal{G}$ is as in (7.3a.1).

The existence of these estimators has been discussed in Dhar (1991a) for $p = 1$ and in Dhar (1991c) for $p \ge 1$. For $p = 1$, these results are relatively easy to state and prove. We give an existence result for the estimator defined at (4) in the case $p = 1$.

Lemma 7.4.1. *In addition to* (7.1.1) *with* $p = 1$, *assume that either*

(5a) $\quad xg(x) \ge 0,\ \forall\ x \in \mathbb{R}, \quad$ *or* $\quad$ (5b) $\quad xg(x) \le 0,\ \forall\ x \in \mathbb{R},$

Then, a minimizer of K_g^+ *exists if either* $G(\mathbb{R}) = \infty$ *or* $G(\mathbb{R}) < \infty$ *and* $g(0) = 0$.

The proof of this lemma is precisely similar to that of Lemma 5.3.1.

The discussion about the computation of their analogues that appears in Section 5.3 is also relevant here with appropriate modifications. Thus, for example, if G is continuous and symmetric around 0, i.e., satisfies (5.3.10), then, analogous to (5.3.7*),

$$\begin{aligned} K_g^+(\mathbf{t}) = \sum_{j=1}^{p}\sum_{i=1}^{n}\sum_{k=1}^{n} g(X_{i-j})g(X_{k-j})\{&|G(X_i - \mathbf{t}'\mathbf{Y}_{i-1}) - G(-X_k + \mathbf{t}'\mathbf{Y}_{k-1})| \\ &- |G(X_i - \mathbf{t}'\mathbf{Y}_{i-1}) - G(X_k - \mathbf{t}'\mathbf{Y}_{k-1})|\}. \end{aligned}$$

If G is degenerate at 0 then one obtains, assuming the continuity of the errors, that

$$K_g^+(\mathbf{t}) = \sum_{j=1}^{p}\Big[\sum_{i=1}^{n} g(X_{i-j})\,\text{sign}(X_i - \mathbf{t}'\mathbf{Y}_{i-1})\Big]^2, \quad \text{w.p.1.} \tag{6}$$

One has similar expressions for a general G. See (5.3.7) and (5.3.7′).

If $g(x) \equiv x \equiv G(x)$, $\hat{\rho}_g$ is m.l.e of ρ if F is logistic, while ρ_g^+ is an analogue of the Hodges–Lehmann estimator. Similarly, if $g(x) \equiv x$ and G is degenerate at 0 then ρ_g^+ is the l.a.d. estimator.

We shall now focus on proving their asymptotic normality. The approach is the same as that of Sections 5.4 and 5.5, i.e., we shall prove that these dispersions satisfy (5.4.A1) – (5.4.A5) by using the techniques that are similar to those used in Section 5.5. Only the tools are somewhat different because of the dependence structure.

To begin with we state the additional assumptions needed under which an asymptotic uniform quadraticity result for a general dispersion of the above type holds. Because here the weights are random, we have to be somewhat careful if we do not wish to impose more than necessary moment conditions on the underlying entities. For the same reason, unlike the linear regression setup where the asymptotic uniform quadraticity of the underlying dispersions was obtained in L_1, we shall obtain these results in probability only. This is also reflected in the formulation of the following assumptions.

(7) (a) $Eh^2(Y_0) < \infty$. (b) $0 < E\epsilon^2 < \infty$.

(8) $\forall\ \|\mathbf{u}\| \le B$, $a \in \mathbb{R}$,

$$\int Eh^2(Y_0)\,|F(x+n^{-1/2}(\mathbf{u}'Y_0+a\|Y_0\|)) - F(x)|\,dG(x) = o(1).$$

(9) There exists a constant $0 < k < \infty$, $\ni\ \forall\ \delta > 0$, $\forall\ \|\mathbf{u}\| \le B$,

$$\liminf_n P\Big(\int n^{-1}\Big[\sum_{i=1}^n h^{\pm}(Y_{i-1})\{F(x+n^{-1/2}\mathbf{u}'Y_{i-1}+n^{-1/2}\delta\|Y_{i-1}\|)- \\ - F(x+n^{-1/2}\mathbf{u}'Y_{i-1}-n^{-1/2}\delta\|Y_{i-1}\|)\}\Big]^2 dG(x) \le k\delta^2\Big) = 1,$$

where $h^{\pm}$ is as in the proof of Theorem 7.2.1.

(10) For every $\|\mathbf{u}\| \le B$,

$$\int n^{-1}\Big[\sum_{i=1}^n h(Y_{i-1})\{F(x+n^{-1/2}\mathbf{u}'Y_{i-1})-F(x)-n^{-1/2}\mathbf{u}'Y_{i-1}f(x)\}\Big]^2 dG(x) = o_p(1),$$

and (5.5.68b) holds.

Now, recall the definitions of W_h, ν_h, $\mathscr{W}_h$, $\mathscr{W}^{\pm}$, $T^{\pm}$, $W^{\pm}$, $Z^{\pm}$, $m^{\pm}$ from (7.1.6), (7.2.2), (7.2.5) and (7.2.6). Let $|\cdot|_G$ denote the L_2-norm w.r.t. the measure G. In the *proofs below*, we have adopted the notation and conventions used in the proof of Theorem 7.2.1. Thus, e.g., $\xi_i \equiv Y_{i-1}$; $\mathscr{W}_{\mathbf{u}}(\cdot)$, $\nu_{\mathbf{u}}(\cdot)$ stand for $\mathscr{W}_h(\cdot, \rho+n^{-1/2}\mathbf{u})$, $\nu_h(\cdot, \rho+n^{-1/2}\mathbf{u})$, etc.

Lemma 7.4.2. *Suppose that the autoregression model* (7.3b.3) *and* (7.3b.4) *holds. Then the following hold.*

(11) *Assumption* (8) *implies that* $\forall\ 0 < B < \infty$,

$$E\int [Z^{\pm}(x; \mathbf{u}, a)-Z^{\pm}(x; \mathbf{u}, 0)]^2 dG(x) = o(1), \quad \forall\ \|\mathbf{u}\| \le B,\ a \in \mathbb{R}.$$

(12) *Assumption* (9) *implies that* $\forall\ 0 < B < \infty$, $\forall\ \|\mathbf{u}\| \le B$,

$$\liminf_n P\Big(\sup_{\|\mathbf{v}-\mathbf{u}\|\le\delta} n^{1/2}|\nu_h^{\pm}(x, \rho+n^{-1/2}\mathbf{v}) - \nu_h^{\pm}(x, \rho+n^{-1/2}\mathbf{u})|_G^2 \le k\delta^2\Big) = 1.$$

where k *and* δ *are as in* (9).

(13) *Assumptions* (7), (9) *and* (10) *imply that* $\forall\ 0 < B < \infty$,

$$\sup_{\|\mathbf{u}\|\le B} \int [n^{1/2}\{\nu_h(x, \rho+n^{-1/2}\mathbf{u}) - \nu_h(x, \rho)\} - \mathbf{u}'n^{-1}\sum_{i=1}^{n} h(Y_{i-1})Y_{i-1}\, f(x)]^2 dG(x) = o_p(1).$$

Proof. Let, for $x, a \in \mathbb{R}$; $\mathbf{u}, \mathbf{y} \in \mathbb{R}^p$,

$$p(x,\mathbf{u},a;\mathbf{y}) := |F(x+n^{-1/2}(\mathbf{u}'\mathbf{y}+a\|\mathbf{y}\|))-F(x+n^{-1/2}\mathbf{u}'\mathbf{y})|. \tag{14}$$

Now, observe that $n^{1/2}[Z^{\pm}(x;\, \mathbf{u},\, a)-Z^{\pm}(x;\, \mathbf{u},\, 0)]$ is a sum of n r.v.'s whose i^{th} summand is conditionally centered, given $\mathcal{F}_{i-1}$, and whose conditional variance, given $\mathcal{F}_{i-1}$, is $E[\{h^{\pm}(\xi_i)\}^2\, p(x,\mathbf{u},a;\xi_i)\{1- p(x,\mathbf{u},a;\xi_i)\}]$, $1 \le i \le n$. Hence, by Fubini, the stationarity of $\{\xi_i\}$ and the fact that $(h^{\pm})^2 \le h^2$, $\forall\ \mathbf{u}\in\mathcal{N}(B)$,

$$\text{l.h.s.}(11) \le \int Eh^2(Y_0)p(x,\mathbf{u},a;Y_0)\, dG(x) = o(1),$$

by (8) applied with the given a and with $a = 0$ and the triangle inequality.

To prove (12), use the nonnegativity of $h^{\pm}$, the monotonicity of F and (7.2.10), to obtain that $\|\mathbf{v}\| \le B$, $\|\mathbf{v} - \mathbf{u}\| \le \delta$ imply that $\forall\ \|\mathbf{u}\| \le B$,

$$n^{1/2}|\nu_{\mathbf{v}}^{\pm}(x) - \nu_{\mathbf{u}}^{\pm}(x)| \le |m^{\pm}(x;\, \mathbf{u},\, \delta) - m^{\pm}(x;\, \mathbf{u}, -\delta)|, \qquad \forall\ x\in\mathbb{R}. \tag{15}$$

This and (9) readily imply (12) as the r.v. in the l.h.s. of (9) is precisely the $|\cdot|_G^2$ of the r.h.s. of (15) for each $n \ge 1$.

The proof of (13) is obtained from (7), (9) and (10) in the same way as that of (5.5.30) from (5.5.7), (5.5.8) and (5.5.9), hence no details are given. □

Lemma 7.4.3. *Suppose that the autoregression model* (7.3b.3) *and* (7.3b.4) *holds. In addition, assume that* (8) *and* (9) *hold.*
Then, $\forall\ 0 < B < \infty$,

$$\text{(16)} \qquad \sup_{\|\mathbf{u}\|\le B} \int [\, \mathscr{W}_h^{\pm}(x, \rho+n^{-1/2}\mathbf{u}) - \mathscr{W}_h^{\pm}(x, \rho)]^2 dG(x) = o_p(1).$$

$$\text{(17)} \qquad \sup_{\|\mathbf{u}\|\le B} \int [\, \mathscr{W}_h(x, \rho+n^{-1/2}\mathbf{u}) - \mathscr{W}_h(x, \rho)]^2 dG(x) = o_p(1).$$

Proof. Let $q(x,\mathbf{u};\mathbf{y}) := |F(x + n^{-1/2}\mathbf{u}'\mathbf{y}) - F(x)|, \ x \in \mathbb{R};\ \mathbf{u}, \mathbf{y} \in \mathbb{R}^p$. The r.v. $n^{1/2}[\, \mathscr{W}_{\mathbf{u}}^{\pm}(\cdot) - \mathscr{W}^{\pm}(\cdot)]$ is a sum of n r.v.'s whose i^{th} summand is conditionally centered, given $\mathcal{F}_{i-1}$, and whose conditional variance, given $\mathcal{F}_{i-1}$, is $E[\{h^{\pm}(\xi_i)\}^2\, q(\cdot,\mathbf{u};\xi_i)\{1- q(\cdot,\mathbf{u};\xi_i)\}]$, $1 \le i \le n$. Hence, by Fubini, the stationarity of $\{\xi_i\}$ and the fact that $(h^{\pm})^2 \le h^2$, $\forall\ \|\mathbf{u}\| \le B$,

$$\text{(18)} \qquad E|\, \mathscr{W}_{\mathbf{u}}^{\pm} - \mathscr{W}^{\pm}|_G^2 \le \int n^{-1} \sum_{i=1}^{n} Eh^2(Y_{i-1})\,|F(x+n^{-1/2}\mathbf{u}'Y_{i-1}) - F(x)|\; dG(x)$$

$$\le \int Eh^2(Y_0)\,|F(x+n^{-1/2}\mathbf{u}'Y_0) - F(x)|\; dG(x).$$

Therefore, by (8) with $a = 0$ and the Markov inequality,

$$\text{(19)} \qquad |\, \mathscr{W}_{\mathbf{u}}^{\pm} - \mathscr{W}^{\pm}|_G^2 = o_p(1), \qquad \forall\ \|\mathbf{u}\| \le B.$$

Thus, to prove (16), because of the compactness of $\mathcal{N}(B)$, it suffices to show that for every $\eta > 0$ there is a $\delta > 0$ such that for every $\|\mathbf{u}\| \le B$,

$$\text{(20)} \qquad \liminf_n P\Big(\sup_{\|\mathbf{v}-\mathbf{u}\|\le\delta} |\mathcal{L}_{\mathbf{v}} - \mathcal{L}_{\mathbf{u}}| < \eta\Big) = 1,$$

where $\mathcal{L}_{\mathbf{u}} := |\, \mathscr{W}_{\mathbf{u}}^{\pm} - \mathscr{W}^{\pm}|_G^2$, $\|\mathbf{u}\| \le B$.

Expand the quadratic, apply the C–S inequality to the cross product terms, to obtain

$$\text{(21)} \qquad |\mathcal{L}_{\mathbf{u}} - \mathcal{L}_{\mathbf{v}}| \le |\, \mathscr{W}_{\mathbf{u}}^{\pm} - \mathscr{W}_{\mathbf{v}}^{\pm}|_G^2 + 2|\, \mathscr{W}_{\mathbf{u}}^{\pm} - \mathscr{W}_{\mathbf{v}}^{\pm}|_G\, |\, \mathscr{W}_{\mathbf{v}}^{\pm} - \mathscr{W}^{\pm}|_G.$$

Observe that $h^{\pm} \ge 0$, F nondecreasing and (7.2.10) imply that

$$0 \le |m^{\pm}(x;\, \mathbf{u}, \pm\delta) - m^{\pm}(x;\, \mathbf{u}, 0)| \le m^{\pm}(x;\, \mathbf{u}, \delta) - m^{\pm}(x;\, \mathbf{u},-\delta),$$

for all $x \in \mathbb{R}$, $\mathbf{s} \in \mathcal{N}(B)$, $\|\mathbf{s}-\mathbf{u}\| \le \delta$. Use this, the second ineqality in (7.2.9), (7.2.10), (7.2.11), and the fact that $(a+b)^2 \le 2(a^2+b^2)$, $a, b \in \mathbb{R}$, to obtain

$$\begin{aligned}|\,\mathscr{W}_{\mathbf{v}}^{\pm}-\mathscr{W}_{\mathbf{u}}^{\pm}\,|_{G}^{2}\le 16\{&\int[Z^{\pm}(x;\mathbf{u},\delta)-Z^{\pm}(x;\mathbf{u},0)]^{2}dG(x)\\&+\int[Z^{\pm}(x;\mathbf{u},-\delta)-Z^{\pm}(x;\mathbf{u},0)]^{2}dG(x)\\&+\int[m^{\pm}(x;\mathbf{u},\delta)-m^{\pm}(x;\mathbf{u},-\delta)]^{2}dG(x)\\&+|n^{1/2}(\nu_{\mathbf{v}}^{\pm}-\nu_{\mathbf{u}}^{\pm})|_{G}^{2}\},\end{aligned}$$

for all $\mathbf{v}\in\mathscr{N}(B)$, $\|\mathbf{v}-\mathbf{u}\|\le\delta$. This together with (9), (12), (13), (19), (21) and the C–S inequality proves (20) and hence, (16).

The proof of (17) follows from (16) and the first inequality in (7.2.9). □

Now define, for $\mathbf{t}\in\mathbb{R}^{p}$,

$$(22)\qquad \mathcal{K}_h(\mathbf{t}) := \int[n^{-1/2}\sum_{i=1}^{n}h(\mathbf{Y}_{i-1})\{I(X_i\le x+\mathbf{t}'\mathbf{Y}_{i-1})-F(y)\}]^{2}\,dG(x),$$

$$\hat{\mathcal{K}}_h(\mathbf{t}) := \int[\,\mathscr{W}_h(x,\rho)+n^{1/2}(\mathbf{t}-\rho)'\,n^{-1}\sum_{i=1}^{n}h(\mathbf{Y}_{i-1})\mathbf{Y}_{i-1}f(x)]^{2}\,dG(x).$$

Theorem 7.4.1. *Suppose that the autoregression model* (7.3b.3) *and* (7.3b.4) *holds and that* (5.5.69), (7) – (10) *hold. Then,* $\forall\ 0<B<\infty$,

$$(23)\qquad \sup_{\|\mathbf{u}\|\le B}|\mathcal{K}_h(\rho+n^{-1/2}\mathbf{u})-\hat{\mathcal{K}}_h(\rho+n^{-1/2}\mathbf{u}))| = o_p(1).$$

Proof. Observe that, by (5.5.69), (7),

$$(24)\qquad E\int\mathscr{W}_h^{2}(x,\rho)\,dG(x) = Eh^{2}(\mathbf{Y}_0)\int F(1-F)dG<\infty.$$

The rest of the proof of (23) follows from Lemmas 7.4.2 and 7.4.3 in a similar way as that of (5.5.28) from Lemmas 5.5.1, 5.5.2 and the result (5.5.30). □

Now we shall apply this result to obtain the required quadraticity of the dispersion K_g and K_g^{+}. For that purpose recall the matrices $\mathscr{X}$, $\mathscr{G}$ and $\mathcal{B}_n$ from (7.3a.1). Note that X_{i-j}, $g(X_{i-j})$ are the $(i,j)^{th}$ entries of $\mathscr{X}$, $\mathscr{G}$, respectively, $1\le i\le n$, $1\le j\le p$. Also observe that the

$$(25)\qquad j^{th}\text{ row of }\mathcal{B}_n\text{ is }\sum_{i=1}^{n}g(X_{i-j})\mathbf{Y}_{i-1}',\quad 1\le j\le p.$$

To obtain the desired result about K_g, we need to apply the above theorem p times, j^{th} time with

$$(26)\qquad h(\mathbf{Y}_{i-1})\equiv g(X_{i-j}),\ j=1,\dots,p.$$

Now write $\mathscr{W}_j$ for $\mathscr{W}_h$ when h is as in (26) and $\mathscr{W}_j(\cdot)$ for $\mathscr{W}_j(\cdot, \rho)$, $1 \le j \le p$. Note that

$$\mathscr{W}_j(x) := n^{-1/2} \sum_{i=1}^{n} g(X_{i-j})\{I(\epsilon_i \le x) - F(x)\}, \qquad 1 \le j \le p, \ \ x \in \mathbb{R}.$$

We also need to define the approximating quadratic forms: For $\mathbf{t} \in \mathbb{R}^p$, let

$$\text{(27)} \qquad \hat{K}_g(\mathbf{t}) := \sum_{j=1}^{p} \int [\, \mathscr{W}_j(x) + n^{1/2}(\mathbf{t}-\rho)'\, n^{-1} \sum_{i=1}^{n} g(X_{i-j})\mathbf{Y}_{i-1} f(x)]^2 dG(x),$$

$$\text{(28)} \qquad \hat{K}_g^{+}(\mathbf{t}) := \sum_{j=1}^{p} \int [\, \mathscr{W}_j^{+}(x) + 2n^{1/2}(\mathbf{t}-\rho)'\, n^{-1} \sum_{i=1}^{n} g(X_{i-j})\mathbf{Y}_{i-1} f(x)]^2 dG(x),$$

where

$$\text{(29)} \qquad \mathscr{W}_j^{+}(x) := n^{-1/2} \sum_{i=1}^{n} g(X_{i-j})\{I(\epsilon_i \le x) - I(-\epsilon_i < x)\}, \qquad 1 \le j \le p, \ x \in \mathbb{R}.$$

Before stating the desired results consider the conditions (7) – (10) when h is as in (26). Condition (7a) is now equal to requiring that $Eg^2(X_{1-j}) < \infty$ for all $j = 1, \dots, p$. Because of the stationarity of $\{X_i\}$, this in turn is equal to

$$\text{(7a}_g\text{)} \qquad Eg^2(X_0) < \infty.$$

Similarly, (8) is equal to

(8_g) $\forall$ $\|\mathbf{u}\| \le B$, $a \in \mathbb{R}$, $1 \le j \le p$,

$$\int Eg^2(X_{1-j})\,|\,F(x+n^{-1/2}(\mathbf{u}'\mathbf{Y}_0 + a\|\mathbf{Y}_0\|)) - F(x)\,|\,dG(x) = o(1).$$

Let (9_g) stand for the condition (9) after $h^{\pm}(\mathbf{Y}_{i-1})$ is replaced by $g^{\pm}(X_{i-j})$, $1 \le j \le p$, in (9), $1 \le i \le n$. Interpret (10_g) similarly. We are now ready to state

Theorem 7.4.2. *Suppose that the autoregression model* (7.3b.3) *and* (7.3b.4) *holds and that* (5.5.68a), (5.5.69), (7b), $(7a_g)$ – (10_g) *hold.*
Then, $\forall$ $0 < B < \infty$,

$$\text{(30)} \qquad \sup_{\|\mathbf{u}\| \le B} |K_g(\rho + n^{-1/2}\mathbf{u}) - \hat{K}_g(\rho + n^{-1/2}\mathbf{u}))| = o_p(1). \qquad \square$$

Proof. Note that the j^{th} summand in K_g is a K_h with h as in (26). Hence (30) readily follows from (23). □

Lemmas 7.4.2 and 7.4.3 can be directly used to obtain the following

Theorem 7.4.3. *In addition to the assumptions of Theorem* 7.4.2, *except* (5.5.69), *assume that* F *is symmetric around* 0, G *satisfies* (5.3.8) *and that* (5.6a.13) *holds.*

Then, $\forall\ 0 < B < \infty$,

(31) $$\sup_{\|\mathbf{u}\|\leq B} |K_g^+(\rho+n^{-1/2}\mathbf{u}) - \hat{K}_g^+(\rho+n^{-1/2}\mathbf{u}))| = o_p(1). \qquad \square$$

Upon expanding the quadratic and using an appropriate analogue of (24) obtained when h is as in (26), one can rewrite

$$\hat{K}_g(\mathbf{t}) = \hat{K}_g(\rho) + 2(\mathbf{t}-\rho)'n^{-1/2}\mathcal{B}_n' \int \mathcal{W}(x)\, f(x)dG(x)$$
$$+ (\mathbf{t}-\rho)'\, n^{-1}\mathcal{B}_n'\mathcal{B}_n\, (\mathbf{t}-\rho)|f|_G^2, \quad \mathbf{t}\in\mathbb{R}^p,$$

where $\mathcal{W} := (\mathcal{W}_1, ..., \mathcal{W}_p)'$. Now consider the r.v.'s in the second term. Recalling the definition of ψ from (5.6a.2), one can rewrite

$$\mathcal{S}_n := \int \mathcal{W}(x)\, f(x)dG(x) = -n^{-1/2}\sum_{i=1}^{n} \mathbf{g}_i\, [\psi(\epsilon_i) - E\psi(\epsilon)],$$

where $\mathbf{g}_i'$ is the i^{th} row of $\mathcal{G}$, i.e.,

(32) $$\mathbf{g}_i' := (g(X_{i-1}), g(X_{i-2}), ..., g(X_{i-p})), \qquad 1 \leq i \leq n.$$

Since $\mathbf{g}_i$ is a function of Y_{i-1}, it is $\mathcal{F}_{i-1}$ — measurable. Therefore, in view of $(7a_g)$ and (5.5.68a), $\{(\mathcal{S}_n, \mathcal{F}_{n-1}),\ n \geq 1\}$ is a mean zero square integrable martingale array. The same assumptions, and an argument like that in the proof of Lemma 7.3d.2, enable one to verify the applicability of Lemma A.3 in the Appendix to $\mathcal{S}_n$. Hence, it follows that

(33) $$\mathcal{S}_n \xrightarrow[d]{} N(0, G^*\tau^2 I_{p\times p}), \quad G^* = E\mathbf{g}_1\mathbf{g}_1', \quad \tau^2 = \mathrm{Var}\, \psi(\epsilon_1)/(\textstyle\int f^2 dG)^2.$$

By the stationarity and the Ergodic Theorem, we also obtain

(34) $$n^{-1}\mathcal{B}_n \longrightarrow \mathcal{B}, \text{ a.s.}, \qquad \mathcal{B} := En^{-1}\mathcal{B}_n = E\mathbf{g}_1\mathbf{Y}_0'.$$

Consequently it follows that the dispersion K_g satisfied (5.4.A1) to (5.4.A3) with $\theta_0 = \rho$, $\delta_n \equiv n^{1/2}$, $S_n \equiv n^{-1/2}\mathcal{B}_n' \mathcal{S}_n$, $W_n \equiv n^{-1}\mathcal{B}_n'\mathcal{B}_n$, $W = \mathcal{B}$, $\Sigma = \mathcal{B}'\, G^*\mathcal{B}\, \tau^2$, and hence it is an u.l.a.n.q. dispersion.

In view of (24) applied to h as in (26), the condition (5.4.A4) is trivially implied by $(7a_g)$ and (5.5.69).

Recall, from Section 5.5, that in the linear regression setup the condition (5.4.A5) was shown to be implied by (5.5.11) and (5.5.12). In the present situation, the role of Γ_n, $\mathbf{\Gamma}_n$ of (5.5.11) is being played by $n^{-1}\mathcal{B}_n$ f, $n^{-1}\mathcal{B}_n \int fdG$, respectively. Thus, in view of (34) and (5.5.68a), an analogue of (5.5.11) would hold in the present case if we additionally assumed that $\mathcal{B}$ is positive definite. An exact analogue of (5.5.12) in the present case is

(35) Either

$$\boldsymbol{\theta}' g_i Y_{i-1}' \boldsymbol{\theta} \geq 0, \ \forall \ 1 \leq i \leq n, \forall \ \boldsymbol{\theta} \in \mathbb{R}^p, \|\boldsymbol{\theta}\| = 1, \ \text{a.s.},$$

or

$$\boldsymbol{\theta}' g_i Y_{i-1}' \boldsymbol{\theta} \leq 0, \ \forall \ 1 \leq i \leq n, \forall \ \boldsymbol{\theta} \in \mathbb{R}^p, \|\boldsymbol{\theta}\| = 1, \ \text{a.s.}.$$

We are now ready to state the following

Theorem 7.4.4. *In addition to the assumptions of Theorem 7.4.2, assume that the $\boldsymbol{B}$ of (34) is positive definite and that (35) holds. Then,*

$$n^{1/2}(\hat{\rho}_g - \rho) = -\{n^{-1}\boldsymbol{B}_n \int f^2 dG\}^{-1} \mathcal{S}_n + o_p(1). \tag{36}$$

Consequently,

$$n^{1/2}(\hat{\rho}_g - \rho) \underset{d}{\longrightarrow} N(0, (\boldsymbol{B})^{-1} \boldsymbol{G}^* (\boldsymbol{B}')^{-1} \tau^2). \tag{37}$$ □

Let $\hat{\rho}_x$ denote the estimator $\hat{\rho}_g$ when $g(x) \equiv x$. Observe that in this case $\boldsymbol{G}^* = \boldsymbol{B} = \mathrm{E}n^{-1}\mathcal{X}'\mathcal{X} = \mathrm{E}Y_0 Y_0'$. Moreover, the assumption (35) is *a priori* satisfied and (7.3b.3), (7.3b.4) and (7b) imply that $\mathrm{E}Y_0 Y_0'$ is positive definite. Consequently, we have obtained

Corollary 7.4.1. *Suppose that the autoregression model* (7.3b.3) *and* (7.3b.4) *holds and that* (5.5.68), (5.5.69), (7b), (8_g) – (10_g) *with* $g(x) \equiv x$ *hold. Then,*

$$n^{1/2}(\hat{\rho}_x - \rho) \underset{d}{\longrightarrow} N(0, (\mathrm{E}Y_0 Y_0')^{-1} \tau^2). \tag{38}$$ □

Remark 7.4.1. *Asymptotic Optimality of $\hat{\rho}_x$.* Because $\boldsymbol{B}$ and $\mathrm{E}Y_0 Y_0'$ are positive definite, and because of $n^{-1}\mathcal{X}'\mathcal{X} \longrightarrow \mathrm{E}Y_0 Y_0'$, a.s., and (34), there exists an N_0 such that $n^{-1}\mathcal{X}'\mathcal{X}$ and $n^{-1}\boldsymbol{B}_n$ are positive definite for all $n \geq N_0$.

Recall the inequality (5.6a.8). Take $\mathbf{J} = n^{-1/2}\mathcal{G}'$, $\mathbf{L} = n^{-1/2}\mathcal{X}'$ in that inequality to obtain

$$n^{-1}\mathcal{G}'\mathcal{G} \geq n^{-1}\mathcal{G}'\mathcal{X}(n^{-1}\mathcal{X}'\mathcal{X})^{-1} \cdot n^{-1}\mathcal{X}'\mathcal{G}, \quad \forall \ n \geq N_0, \ \text{a.s.},$$

with equality holding if, and only if $\mathcal{X} \propto \mathcal{G}$. Letting n tend to infinity in this inequality yields

$$(\boldsymbol{B})^{-1} \boldsymbol{G}^* (\boldsymbol{B}')^{-1} \geq (\mathrm{E}Y_0 Y_0')^{-1}.$$

We thus have proved the following:

(39) *Among all estimators* $\{\hat{\rho}_g;$ g *satifying* ($7a_g$) – (10_g) *for the given* (F, G) *that satisfy* (7b), (5.5.68), (5.5.69)}, *the one that minimizes the asymptotic variance* is $\hat{\rho}_x$! □

We shall now state analogous results for ρ_g^+. Arguments for their proofs are similar to those appearing above and, hence, will not be given.

Theorem 7.4.5. *In addition to the assumptions of Theorem* 7.4.4, *except* (5.5.69), *assume that* F *is symmetric around* 0, G *satisfies* (5.3.8) *and that* (5.6a.13) *holds. Then,*

$$n^{1/2}(\rho_g^+ - \rho) = -\{n^{-1}\mathcal{B}_n \int f^2 dG\}^{-1}\, \mathcal{S}_n^{\,+} + o_p(1), \tag{40}$$

where

$$\mathcal{S}_n^{\,+} := \int \mathcal{W}^+(x)\, f(x)dG(x) = n^{-1/2}\sum_{i=1}^{n} g_i\, [\psi(-\epsilon_i) - \psi(\epsilon_i)].$$

Consequently,

$$n^{1/2}(\rho_g^+ - \rho) \xrightarrow[d]{} N(0, (\mathcal{B})^{-1}\mathcal{G}^*(\mathcal{B}')^{-1}\tau^2), \tag{41}$$

$$n^{1/2}(\rho_x^+ - \rho) \xrightarrow[d]{} N(0, (EY_0Y_0')^{-1}\tau^2). \tag{42}$$ □

Obviously the optimality property like (39) holds here also.

Remark 7.4.2. *On assumptions for the asymptotic normality of* $\hat{\rho}_x$, ρ_x^+. If G is a finite measure and F has uniformly continuous density then it is not hard to see that (8_g) – (10g), with g(x) ≡ x, are all implied by (7b).

Consider the following assumptions for general G:

(43) $E|\epsilon|^3 < \infty, \quad E\epsilon^2 > 0.$

(44) As a function of $s \in \mathbb{R}$, $\int E|X_{1-j}|^2\|Y_0\|\, f(x+s\|Y_0\|)\, dG(x)$ is continuous at 0, $1 \le j \le p$.

(45) For every $\delta > 0$, $\mathbf{u} \in \mathbb{R}^p$,

$$\int_{-1}^{1}\int E\{\|Y_0\|[f(x+n^{-1/2}(\mathbf{u}'Y_0+t\delta\|Y_0\|)) - f(x+n^{-1/2}\mathbf{u}'Y_0)]\}^2\, dG(x)dt = o(1).$$

(46) For every $\mathbf{u} \in \mathbb{R}^p$,

$$\int [n^{-1} \sum_{i=1}^{n} X_{i-j}^{\pm} \|Y_{i-1}\| \, f(x+n^{-1/2}\mathbf{u}'Y_{i-1})]^2 \, dG(x) = O_p(1), \quad 1 \le j \le p.$$

An argument similar to the one used in verifying the Claim 5.5.1 shows that (5.5.68a), (43) and (44) imply (8_g) while (5.5.68b), (45) and (46) imply (9g) and (10g).

In particular if $G(x) \equiv x$, then (5.5.68), (43) and f continuous imply all of the above conditions, (5.5.69) and (5.6a.13). This is seen with the help of a version of Scheffe's Theorem. □

Remark 7.4.3. *Asymptotic relative efficiency of* $\hat{\rho}_x$, ρ_x^+. Since their asymptotic variances are the same, we shall carry out the discussion in terms of $\hat{\rho}_x$ only, as the same applies to ρ_x^+ under the additional assumption of the symmetry of F and G.

Consider the case $p = 1$. Let $\sigma^2 = \mathrm{Var}(\epsilon)$ and $\hat{\rho}_{ls}$ denote the least square estimator of ρ_1. Then it is well known that under (7b), $n^{1/2}(\hat{\rho}_{ls} - \rho_1) \xrightarrow[d]{} N(0, 1-\rho_1^2)$. See, e.g., Anderson (1971). Also note that in this case $(EY_0Y_0')^{-1} = (1-\rho_1^2)/\sigma^2$. Hence the asymptotic relative efficiency e of $\hat{\rho}_x$, relative to $\hat{\rho}_{ls}$, obtained by taking the ratio of the inverses of their asymptotic variances, is

$$e = e(\hat{\rho}_x, \hat{\rho}_{ls}) = \sigma^2/\tau^2. \tag{47}$$

Note that $e > 1$ means $\hat{\rho}_x$ is asymptotically more efficient than $\hat{\rho}_{ls}$. It follows that $\hat{\rho}_x$ is to be prefered to $\hat{\rho}_{ls}$ for the heavy tailed error d.f.'s F. Also note that if $G(x) \equiv x$ then $\tau^2 = 1/12\,[\int f^2(x)\,dx]^2$ and $e = 12\,\sigma^2\,[\int f^2(x)\,dx]^2$. If G is degenerate at 0, then $\tau^2 = 1/[4f^2(0)]$ and $e = 4\sigma^2 f^2(0)$. These expressions are well known in connection with the Wilcoxon and median rank estimators of the slope parameters in linear regression models. For example if F is $N(0, 1)$ then the first expression is $3/\pi$ while the second is $2/\pi$. See Lehmann (1975) for some bounds on these expressions. Similar conclusions remain valid for $p > 1$. □

Remark 7.4.4. *Least Absolute Deviation Estimator.* As mentioned earlier, if we choose $g(x) \equiv x$ and G to be degenerate at 0 then ρ_x^+ is the l.a.d. estimator, v.i.z.,

$$\rho_{lad}^+ := \mathrm{argmin}\{\sum_{j=1}^{p}[\sum_{i=1}^{n} X_{i-j}\,\mathrm{sign}(X_i - \mathbf{t}'Y_{i-1})]^2;\ \mathbf{t} \in \mathbb{R}^p\} \tag{48}$$

See also (6). Because of its importance we will now summarize sufficient conditions under which it is asymptotically normally distributed. Of course we could use the stronger conditions (43) – (46) but they do not use the given information about G.

Clearly, (7b) implies ($7a_g$) when $g(x) \equiv x$. Moreover, in this case the l.h.s. of (8_g) is

$$EX_{1-j}^2 |F(n^{-1/2}(\mathbf{u}'\mathbf{Y}_0 + a\|\mathbf{Y}_0\|)) - F(n^{-1/2}\mathbf{u}'\mathbf{Y}_0)|$$

which tends to 0 by the D.C.T., (7b) and the continuity of F, $1 \le j \le p$.

Now consider (9_g). Asssume the following:

(49) F *has a density* f *that is continuous at* 0 *and* $f(0) > 0$.

Recall from (7.3b.6) that under (7.3b.3), (7.3b.4) and (7b),

$$n^{-1/2}\max\{\|\mathbf{Y}_{i-1}\|;\ 1 \le i \le n\} = o_p(1). \tag{50}$$

The r.v.'s involved in the l.h.s. of (9_g) in the present case are

$$n^{-1}[\sum_{i=1}^{n} X_{i-j}^{\pm}\{F(n^{-1/2}\mathbf{u}'\mathbf{Y}_{i-1} + n^{-1/2}\delta\|\mathbf{Y}_{i-1}\|) \\ - F(n^{-1/2}\mathbf{u}'\mathbf{Y}_{i-1} - n^{-1/2}\delta\|\mathbf{Y}_{i-1}\|)\}]^2$$

which, in view of (49), can be bounded above by

$$4\delta^2 [n^{-1}\sum_{i=1}^{n} X_{i-j}^{\pm}\|\mathbf{Y}_{i-1}\|\, f(\eta_{ni})]^2, \tag{51}$$

where $\{\eta_{ni}\}$ are r.v.'s, $\eta_{ni} \in n^{-1/2}[\mathbf{u}'\mathbf{Y}_{i-1} - \delta\|\mathbf{Y}_{i-1}\|,\ \mathbf{u}'\mathbf{Y}_{i-1} + \delta\|\mathbf{Y}_{i-1}\|]$, $1 \le i \le n$. Hence, by the stationarity and the ergodicity of the process $\{X_i\}$, (7b), (49) and (50) imply that the r.v.'s in (51) converge to $4\delta^2\,[EX_{1-j}^{\pm}\|\mathbf{Y}_0\|\ f(0)]^2$, a.s., $1 \le j \le p$. This verifies (9_g) in the present case. One similarly verifies (10_g).

Also note that here (5.5.68) is implied by (49) and (5.5.69) is trivially satisfied as $\int F(1-F)\, dG \le 1/4$ in the present case. We summarize the above discussion in

Corollary 7.4.3. *Assume that the autoregression model* (7.3b.3) *and* (7.3b.4) *holds. In addition, assume that the error d.f.* F *has finite second moment,* $F(0) = 1/2$ *and satisfies* (49). *Then,*

$$n^{1/2}(\rho_{\ell ad}^{+} - \rho) \xrightarrow[d]{} N(0, (E\mathbf{Y}_0\mathbf{Y}_0')^{-1}/4f^2(0)),$$

where $\rho_{\ell ad}^{+}$ *is defined at* (48). □

7.5. GOODNESS-OF-FIT TESTING.

Once again consider the AR(p) model given by (7.3b.3), (7.3b.4) and let F_0 be a known d.f.. Consider the problem of testing H_0: $F = F_0$. One of the common tests of H_0 is based on the Kolmogorov–Smirnov statistic

$$D_n := n^{1/2}\sup_x |F_n(x, \hat{\rho}) - F_0(x)|.$$

From Corollary 7.2.1 one readily has the following:

If F_0 *has a uniformly continuous density* f_0, $f_0 > 0$ a.e.; $\int x^2 dF_0(x) < \infty$, $\hat{\rho}$ *satisfies* (7.3c.(iv)) *under* F_0, *then, under* H_0,

$$D_n = \sup |B(F_0(x)) + n^{1/2}(\hat{\rho} - \rho)' n^{-1}\Sigma_i Y_{i-1} f_0(x)| + o_p(1).$$

In addition, if $EY_0 = 0 = E\epsilon_1$, *then* $D_n \underset{d}{\longrightarrow} \sup\{|B(t)|, 0 \le t \le 1\}$, *thereby rendering* D_n *asymptotically distribution free.*

Next, consider, H_{01}: $F = N(\mu, \sigma^2)$, $\mu \in \mathbb{R}$, $\sigma^2 > 0$. In other words, H_{01} states that the AR(p) process is generated by some normal errors. Let $\hat{\mu}_n$, $\hat{\sigma}_n$ and $\hat{\rho}_n$ be estimators of μ, σ, ρ respectively. Define

$$\hat{F}_n(x) := n^{-1} \Sigma_i I(X_i \le x\hat{\sigma}_n + \hat{\mu}_n + \hat{\rho}_n' Y_i), \qquad x \in \mathbb{R},$$

$$\hat{D}_n := n^{1/2} \sup_x |\hat{F}_n(x) - \Phi(x)|, \quad \Phi = N(0, 1) \text{ d.f.}.$$

Corollary 7.2.1. can be readily modified in a routine fashion to yield that if

$$n^{1/2}|(\hat{\mu}_n - \mu) + (\hat{\sigma}_n - \sigma)| \sigma^{-1} + n^{1/2}\|\hat{\rho}_n - \rho\| = O_p(1)$$

then

$$\hat{D}_n := \sup_x| B(\Phi(x)) + n^{1/2}\{(\hat{\mu}_n - \mu) + (\hat{\sigma}_n - \sigma)\}\sigma^{-1} n(x)| + o_p(1),$$

where n is the density of Φ. Thus the asymptotic null distribution of $\hat{D}_n$ is similar to its analogue in the one sample location–scale model: *the estimation of* ρ *has no effect on the large sample null distribution of* $\hat{D}_n$.

Clearly, similar conclusions can be applied to other goodness-of-fit tests. In particular we leave it as *an exercise* for an interested reader to investigate the large sample behaviour of the goodness-of-fit tests based on L_2–distances, analogous to the results obtained in Section 6.3. Lemma 6.3.1 and the results of the previous section are found useful here. □□□□□

APPENDIX

We include here some results relevant to the weak convergence of processes in $\mathbb{D}[0, 1]$ and $\mathbb{C}[0, 1]$ for the sake of easy reference and without proofs. Our source is the book by Billingsley (1968) (B) on *Convergence of Probability Measures.*

To begin with, let $\xi_1, \ldots, \xi_m$ be r.v.'s, not necessarily independent and define

$$S_k := \sum_{j=1}^{k} \xi_j, \quad 1 \le k \le m; \qquad M_m := \max_{1 \le k \le m} |S_k|.$$

The following lemma is obtained by combining (12.5), (12.10) and Theorem 12.1 from pp 87–89 of (B).

Lemma A.1. *Suppose there exist nonnegative numbers* $u_1, u_2, \ldots, u_m$, *a* $\gamma \ge 0$ *and an* $\alpha > 0$ *such that*

$$E\{|S_k - S_j|^{\gamma}|S_j - S_i|^{\gamma}\} \le \left(\sum_{r=i+1}^{k} u_r\right)^{2\alpha}, \qquad 0 \le i \le j \le k \le m.$$

Then, $\forall\ \lambda > 0$,

$$P(M_m \ge \lambda) \le K_{\gamma,\alpha} \cdot \lambda^{-2\gamma}\left(\sum_{r=1}^{m} u_r\right)^{2\alpha} + P\left(|S_m| \ge \frac{\lambda}{2}\right),$$

where $K_{\gamma,\alpha}$ *is a constant depending only on* γ *and* α.

The following inequality is given as Corollary 8.3 in (B).

Lemma A.2. *Let* $\{\zeta(t),\ 0 \le t \le 1\}$ *be a stochastic process on some probability space. Let* $\delta > 0$, $0 = t_0 < t_1 < \ldots < t_r = 1$ *with* $t_i - t_{i-1} \ge \delta$, $2 \le i \le r-1$ *be a partition of* $[0, 1]$. *Then,* $\forall\ \epsilon > 0, \forall\ 0 < \delta \le 1$,

$$P\left(\sup_{|t-s|<\delta} |\zeta(t) - \zeta(s)| \ge 3\epsilon\right) \le \sum_{i=1}^{r} P\left(\sup_{t_{i-1} \le t \le t_i} |\zeta(t) - \zeta(t_{i-1})| \ge \epsilon\right).$$

Definition: A sequence of stochastic processes $\{\zeta_n\}$ in $\mathbb{D}[0, 1]$ is said to converge weakly to a stochastic process $\zeta \in \mathbb{C}[0, 1]$ if every finite dimensional distribution of $\{\zeta_n\}$ converges weakly to that of ζ and if $\{\zeta_n\}$ is tight with respect to the uniform metric.

The following theorem gives sufficient conditions for the weak convergence of a sequence of stochastic processes in $\mathbb{D}[0, 1]$ to a limit in $\mathbb{C}[0, 1]$. It is essentially Theorem 15.5, p 127 of (B).

Theorem A.1. *Let* $\{\zeta_n(t),\ 0 \le t \le 1\}$ *be a sequence of stochastic processes in* $\mathbb{D}[0, 1]$. *Suppose that* $|\zeta_n(0)| = O_p(1)$ *and that* $\forall\ \epsilon > 0$,

$$\lim_{\eta \to 0} \limsup_n P\Big(\sup_{|s-t|<\eta} |\zeta_n(s) - \zeta_n(t)| \ge \epsilon\Big) = 0.$$

Then the sequence $\{\zeta_n(t),\ 0 \le t \le 1\}$ *is tight, and of* ζ *is the weak limit of a subsequence* $\{\zeta_{n'}(t),\ 0 \le t \le 1\}$, *then* $P(\zeta \in \mathbb{C}[0, 1]) = 1$.

The following theorem gives sufficient conditions for the weak convergence of a sequence of stochastic processes in $\mathbb{C}[0, 1]$ to a limit in $\mathbb{C}[0, 1]$. It is essentially Theorem 12.3, p 95 of (B).

Theorem A.2. *Let* $\{\zeta_n(t),\ 0 \le t \le 1\}$ *be a sequence of stochastic processes in* $\mathbb{C}[0, 1]$. *Suppose that* $|\zeta_n(0)| = O_p(1)$ *and that there exist a* $\gamma \ge 0$, $\alpha > 1$ *and a nondecreasing continuous function* F *on* $[0, 1]$ *such that,*

$$P(|\zeta_n(t) - \zeta_n(s)| \ge \lambda) \le \lambda^{-\gamma} |F(t) - F(s)|^{\alpha}$$

holds for all s, t *in* [0,1] *and for all* $\lambda > 0$.

Then the sequence $\{\zeta_n(t),\ 0 \le t \le 1\}$ *is tight, and if* ζ *is the weak limit of a subsequence* $\{\zeta_{m_n}(t),\ 0 \le t \le 1\}$, *then* $P(\zeta \in \mathbb{C}[0, 1]) = 1$.

We also need a central limit theorem for martingale arrays. Let $(\Omega, \mathcal{F}, P)$ be a probability space; $\{\mathcal{F}_{n,i},\ 1 \le i \le n\}$, be an array of sub σ–fields such that $\mathcal{F}_{n,i} \subset \mathcal{F}_{n,i+1}$, $1 \le i \le n$; X_{ni} be $\mathcal{F}_{n,i}$ measurable r.v. with $EX_{ni}^2 < \infty$, $E(X_{ni}|\ \mathcal{F}_{n,i-1}) = 0$, $1 \le i \le n$; and let $S_{nj} = \Sigma_{i\le j} X_{ni}$, $1 \le j \le n$. Then $\{S_{ni}, \mathcal{F}_{n,i};\ 1 \le i \le n,\ n \ge 1\}$ is called *a zero–mean square–integrable martingale array with differences* $\{X_{ni};\ 1 \le i \le n,\ n \ge 1\}$.

The central limit theorem we find useful is Corollary 3.1 of Hall and Heyde (1980) which we state here for an easy reference.

Lemma A.3. *Let* $\{S_{ni}, \mathcal{F}_{n,i};\ 1 \le i \le n,\ n \ge 1\}$ *be a zero–mean square–integrable martingale array with differences* $\{X_{ni}\}$ *satisfying the following conditions.*

(1) $\forall\ \epsilon > 0$, $\quad \sum_{i=1}^{n} E[X_{ni}^2 I(|X_{ni}| > \epsilon)|\mathcal{F}_{n,i-1}] = o_p(1)$.

(2) $\quad \sum_{i=1}^{n} E[X_{ni}^2|\mathcal{F}_{n,i-1}] \longrightarrow$ a r.v. η^2, in probability.

(3) $\quad \mathcal{F}_{n,i} \subset \mathcal{F}_{n+1,i}$, $\ 1 \le i \le n$, $\ n \ge 1$.

Then S_{nn} *converges in distribution to a r.v.* Z *whose characteristic function at* t *is* $E\exp(-\eta^2 t^2/2)$, $t \in \mathbb{R}$. □

BIBLIOGRAPHY

Adichie, J. (1967). Estimation of regression parameters based on rank tests. *Ann. Math. Statist.* **38**, 894–904.

Anderson, T.W. (1971). *The statistical analysis of time series.* J. Wiley & Sons, Inc. New York.

Anderson, T.W. and Darling, D.A. (1952). Asymptotic theory of certain 'goodness of fit' criteria based on stochastic processes. *Ann. Math. Statist.* **23**, 193–212.

Babu, J. and Singh, K. (1983). Inference on means using bootstrap. *Ann. Statist.* **11**, 999–1003.

_________ (1984). On one term Edgeworth correction by Efron's bootstrap. *Sankhya, Series. A, Pt. 2.* **46**, 219–232.

Basawa, I.V. and Koul, H.L. (1988). Large sample statistics based on quadratic dispersions. *Int. Statist. Review* **56**, 199–219.

Beran, R.J. (1977). Minimum Hellinger distance estimates for parametric models. *Ann. Statist.* **5**, 445–463.

_________ (1978). An efficient and robust adaptive estimator of location. *Ann. Statist.* **6**, 292–313.

_________ (1982). Robust estimation in models for independent non–identically distributed data. *Ann. Statist.* **2**, 415–428.

Bickel, P.J. (1982). On adaptive estimation. *Ann. Statist.* **10**, 647–671.

Bickel, P.J. and Freedman, D. (1981). Some asymptotic theory for the bootstrap. *Ann. Statist.* **2**, 415–428.

Bickel, P.J. and Lehmann, E.L. (1976). Descriptive statistics for nonparametric models. III Dispersion. *Ann. Statist.* **4**, 1139–1158.

Bickel, P.J. and Wichura, M. (1971). Convergence criteria for multiparameter stochastic processes and some applications. *Ann. Math. Statist.* **42**, 1656–1670.

Billingsley, P. (1968). *Convergence of Probability Measures.* John Wiley and Sons, New York.

Brockwell, P. J. and Davis, R. A. (1987). *Time Series: Theory and Methods.* Springer–Verlag, New York.

Brown, L.D. and Purves, R. (1973). Measurable selections of extrema. *Ann. Statist.* **1**, 902–912.

Boldin, M. V. (1982). Estimation of the distribution of noise in autoregression scheme. *Theor. Probab. Appl.* **27**, 866–871.

_________ (1989). On testing hypotheses in the sliding average scheme by the Kolmogorov–Smirnov and ω^2 tests. *Theor. Probab. Appl.* **34**, 699–704.

Boos, D.D. (1981). Minimum distance estimators for location and goodness of fit. *J. Amer. Statist. Assoc.* **76**, 663–670.

_________ (1982a). Minimum Anderson–Darling estimation. *Comm. Stat. Theor. Meth.* **11** (24), 2747–2774.

_________ (1982b). A test of symmetry associated with the Hodges–Lehmann estimator. *J. Amer. Statist. Assoc.* **77**, 647–651.

Bustos, O. H. (1982). General M-estimates for contaminated pth order autoregressive processes: consistency and asymptotic normality. *Zeit fur Wahrscheinlichkeitstheorie*, **59**, 491–504.

Cheng, K.F. and Serfling, R.J. (1981). On estimation of a class of efficiency–related parameters. *Scand. Act. J.* 83–92.

Dehling, H. and Taqqu, M. S. (1989). The empirical process of some long range dependent sequences with an application to U–statistics. *Ann. Statist.* **17**, 1767–1783.

Delong, D. (1983). Personal Communication. S.A.S. Institute, Inc. Cary. N.C., 27511–8000.

Denby, L. and Martin, D. (1979). Robust estimation of the first order autoregressive parameter. *J. Amer. Statist. Assoc.* **74**, 140–146.

Dhar, S. K. (1991a). Minimum distance estimation in an additive effects outliers model. *Ann. Statist.* **19**, 205–228.

_________ (1991b). Computation of minimum distance estimators in multiple linear regression model. *Comm. Statist. Theor. and Meth. Ser. B*, **21**(1), (1992), 18pps.

_________ (1991c). Existence and compulation of minimum distance estimators in AR(k) model. To appear in *J.A.S.A.*, March 1993.

Donoho, D. L. and Liu, R. C. (1988). The " automatic" robustness of minimum distance functionals. *Ann. Statist.* **16**, 552–586.

_________ (1988). Pathologies of some minimum distance estimators. *Ann. Statist.* **16**, 587–608.

Dupač, V. and Hájek, J. (1969). Asymptotic normality of linear rank statistics II. *Ann. Math. Statist.* **40**, 1992–2017.

Durbin, J. (1973). *Distribution theory for tests based on the sample d.f.* Philadelphia: SIAM.

_________ (1976). Kolomogorov–Smirnov tests when parameters are estimated. *Empirical distributions and Processes*; *Springer–Verlag Lecture Notes in Math.*, #566.

Dvortzky, A., Kiefer, J., and Wolfowitz, J. (1956). Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *Ann. Math. Statist.* **27**, 642–669.

Efron, B. (1979). Bootstrap methods: another look at the jackknife. *Ann. Statist.* **7**, 1–26.

_________ (1982). *The Jackknife, the Bootstrap and Other resampling plans.* CBMS–NSF Reg. Conf. Ser. **38**. S. I. A. M.

Einmahl, U. and Mason, D. (1992). Approximations to permutation and exchangeable processes. *J. Theor. Probab.* **5**, 101–126.

Eyster, J. (1977). Asymptotic normality of simple linear random rank statistics under the alternatives. Ph. D. thesis, Michigan State University, East Lansing.

Fabian, V and Hannan, J. (1982). On estimation and adaptive estimation for locally asymptotically normal families. *Z. Wahrsch. Verw. Gebiete* **59** 459–478.

Feller, W. (1966). *An Introduction to Probability Theory and its Applications* **II**. John Wiley & Sons, New York.

Fine, T. (1966). On Hodges–Lehmann shift estimator in the two sample problem. *Ann. Math. Statist.* **37**, 1814–1818.
Finkelstein, H. (1971). The law of the iterated logarithm for empirical distributions. *Ann. Math. Statist.* **42**, 607–615.
Freedman, D. (1981). Bootstrapping regression models. *Ann. Statist.* **9**, 1218–1228.
Ghosh, M. and Sen, P.K. (1972). On bounded confidence intervals for regression coefficients based on a class of rank statistics. *Sankhya* A, **34**, 33–52.
Hall, P. and Heyde, C. C. (1980). *Martingale limit theory and its applications.* Academic Press. New York.
Hampel, F. (1971). A general qualitative definition of robustness. *Ann. Math. Statist.* **42**, 1887–1896.
Hájek, J. and Sĭdák, Z. (1967). *Theory of rank tests.* Academic Press. New York.
Hájek, J. (1969). *Nonparametric Statistics.* Holden Day. San Francisco.
_______ (1972). Local asymptotic minimax and admissibility in estimation. *Proc. Sixth. Berkeley Symp. Math. Statist. Probability.* **1**, 597–616. University of California Press.
Hardy, G. H., Littlewood, J. E. and Polya, G. (1952). *Inequalities* (2nd Ed.). Cambridge University Press.
Heiler, S. and Willers, R. (1988). Asymptotic normality of R–estimates in the linear model. *Statistics*, **19**, 173–184.
Hettmansperger, T. P. (1984). *Statistical inference based on ranks.* J. Wiley & Sons. Inc. New York.
Hodges, J.L., Jr. and Lehmann, E.L. (1963). Estimates of location based on rank tests. *Ann. Math. Statist.* **34**, 598–611.
Huber, P. (1973). Robust regression: Asymptotics, conjectures and Monte Carlo. *Ann. Statist.* **1**, 799–821.
Huber, P. (1981). *Robust Statistics.* Wiley, New York.
Johnson, W. B., Schechtman, G. and Zinn, J. (1985). Best constants in moment inequalities for linear combinations of independent and exchangeable random variables. *Ann. Probab.* **13**, 234–253.
Jaeckel, L.A. (1972). Estimating regression coefficient by minimizing the dispersion of residuals. *Ann. Math. Statist.* **43**, 1449–1458.
Jurečková, J. (1969). Asymptotic linearity of rank test statistics in regression parameters. *Ann. Math. Statist.* **40**, 1889–1900.
___________ (1971). Nonparametric estimates of regression coefficients. *Ann. Math. Statist.* 42, 1328–1338.
Kac, M., Kiefer, J. and Wolfowitz, J. (1955). On tests of normality and other tests of goodness-of-fit based on distance methods. *Ann. Math. Statist.* **26**, 189–211.
Kiefer, J. (1959). K-sample analogues of the Kolmogorov–Smirnov and Cramer–von Mises tests. *Ann. Math. Statist.* **30**, 420–447.
Koul, H.L. (1969). Asymptotic behavior of Wilcoxon type confidence regions in multiple linear regression. *Ann. Math. Statist.* **40**, 1950–1979.

Koul, H.L. (1970). Some convergence theorems for ranks and weighted empirical cumulatives. *Ann. Math. Statist.* **41**, 1768–1773.

________ (1971). Asymptotic behavior of a class of confidence regions based on rank in regression. *Ann. Math. Statist.* **42**, 466–476.

________ (1977). Behavior of robust estimators in the regression model with dependent errors. *Ann. Statist.* **5**, 681–699.

________ (1979). Weighted empirical processes and the regression model. *J. Indian Statist. Assoc.* **17**, 83–91.

________ (1980). Some weighted empirical inferential procedures for a simple regression model. *Colloq. Math. Soc. Janos Bolyai*, **32**. *Nonpar. Statist. Inf.* (537–565).

________ (1984). Tests of goodness-of-fit in linear regression. *Colloq. Math. Soc. Janos Bolyai*, **45**. *Goodness–of–Fit.* (279–315).

________ (1985a). Minimum distance estimation in multiple linear regression. *Sankhya, Ser. A*, **47**, **Part 1**, 57–74.

________ (1985b). Minimum distance estimation in linear regression with unknown errors. *Statist. & Prob. Letters.* 3, 1–8.

________ (1986). Minimum distance estimation and goodness of fit in first–order autoregression. *Ann. Statist.* **14** 1194–1213.

________ (1989). A quadraticity limit theorem useful in linear models. *Probab. Th. Rel. Fields* **82** 371–386.

________ (1991). A weak convergence result useful in robust autoregression. *J. Statist. Planning and Inference*, **29**, 291–308.

Koul, H.L. and DeWet, T. (1983). Minimum distance estimation in a linear regression model. *Ann. Statist.* **11**, 921–932.

Koul, H.L. and Levental, S. (1989). Weak convergence of residual empirical processes in explosive autoregression. *Ann. Statist.* **17**, 1784–1794.

Koul, H.L. and Mukherjee, K. (1992). Asymptotics of R–, MD– and LAD–estimators in linear regression models with long range dependent errors. A preprint.

Koul, H.L. and Ossiander, M. (1992). Weak convergence of randomly weighted dependent residual empiricals with application to autoregression. RM # 520, Department of Statist. & Probab., MSU, East Lansing.

Koul, H.L., Sievers, G. L., and McKean, J. W. (1987). An estimator of the scale parameter for the rank analysis of linear models under general score functions. *Scand. J. Statist.* **14**, 131–141.

Koul, H.L. and Staudte, R.G., Jr. (1972). Weak convergence of weighted empirical cumulatives based on ranks. *Ann. Math. Statist.* **43**, 832–841.

Koul, H.L. and Susarla, V. (1983). Adaptive estimation in linear regression. *Statist. and Decis.* **1**, 379–400.

Koul, H.L. and Zhu, Z. (1991). Bahadur representations for some minimum distance estimators in linear models. A preprint.

Kreiss, P. (1991). Estimation of the distribution function of noise in stationary processes. *Metrika*, **38**, 285–297.

Kuelbs, J. (1976). A strong convergence theorem for Banach spaces valued r.v.'s *Ann. Prob.* **4**, 744–771.

Lahiri, S. (1989). Bootstrap approximations to the distributions of M-estimators. Ph. D. thesis. Michigan State Univ., East Lansing.

________ (1991). Bootstrapping M-estimators of a multiple linear regression parameter. A preprint. To appear in the *Ann. Statist.*

Le Cam, L. (1972). Limits of experiments. *Proc. Sixth. Berkeley Symp. Math. Statist. Probability.* 1, 245–261. University of California Press.

________ (1986). *Asymptotical methods in statistical theory.* Springer, New York.

Lehmann, E. L. (1963). Nonparametric confidence intervals for a shift parameter. *Ann. Math. Statist.* **34**, 1507–1512.

________ (1975). *Nonparametrics.* Holden Day, San Francisco.

Lehmann, E. L. (1986). *Testing Statistical Hypotheses,* 2nd Edition. J. Wiley & Sons, New York.

Levental, S. (1989). A uniform CLT for uniformly bounded families of martingale–differences. *J. Theoret. Probab.* **2**, 271–287.

Loève, M. (1963). *Probability Theory.* D. Van Nostrand Co., Inc. Princeton, New Jersey.

Loynes, R.M. (1980). The empirical d.f. of residuals from generalized regression. *Ann. Statist.* **8**, 285–298.

Marcus, M.B. and Zinn, J. (1984). The bounded law of iterated logarithm for the weighted empirical distribution process in the non i.i.d. case. *Ann. Prob.* **12**, 335–360.

Martynov, G.V. (1975). Computation of distribution functions of quadratic forms of normally distributed r.v.'s *Theor. Probab. Appl.* **20**, 782–793.

________ (1976). Computation of limit distributions of statistics for normality tests of type ω^2. *Theor. Probab. Appl.* **21**, 1–13.

Mehra, K.L. and Rao, M.S. (1975). Weak convergence of generalized empirical processes relative to d_q under strong mixing. *Ann. Prob.* **3**, 979–991.

Millar, P.W. (1981). Robust estimation via minimum distance methods. *Zeit fur Wahrscheinlichkeitstheorie,* **55**, 73–89.

________ (1982). Optimal estimation of a general regression function. *Ann. Statist.* **10**, 717–740.

________ (1984). A general approach to the optimality of minimum distance estimators. *Trans. Amer. Math. Soc.* **286**, 377–418.

Noether, G. E. (1949). On a theorem by Wald and Wolfowitz. *Ann. Math. Statist.* **20**. 455–458.

Parr, W. (1981). Minimum distance estimation: a bibliography. *Comm. Statist. Theor. Meth.* **A10** (12), 1205–1224.

Parr, W.C. and Schucany, W.R. (1979). Minimum distance and robust estimation. *J.A.S.A,* **75**, 616–624.

Parthasarathy, K.R. (1967). *Probability measures on metric spaces.* Academic Press, New York.

Pollard, D. (1984). *Convergence of Stochastic Processes.* Springer, N.Y.

________ (1991). Asymptotics for least absolute deviation regression estimators. *Econometric Theor.* **7**, 186–199.

Puri, M.L. and Sen, P.K. (1969). *Nonparametric methods in Multivariate Analysis.* John Wiley and Sons, New York.

Rao, K.C. (1972). The Kolmogoroff, Cramer–von Mises Chisquares statistics for goodness-of-fit tests in the parametric case. (abstract). *Bull. Inst. Math. Statist.* **1**, 87.

Rao, P. V., Schuster, E., and Littel. R. (1975). Estimation of shift and center of symmetry based on Kolmogorov–Smirnov statistics. *Ann Statist.* **3**, 862–873.

Rockafeller, R. T. (1970). *Convex Analysis.* Princeton University Press, Princeton, N.J.

Schweder, T. (1975). Window estimation of the asymptotic variance of rank estimators of location. *Scand. J. Statist.* **2**, 113–126.

Shorack, G. (1973). Convergence of reduced empiricals and quantile processes with applications to functions of order statistics in non-i.i.d. case. *Ann Statist.* **1**, 146–152.

_________ (1979). Weighted empirical process of row independent r.v.'s with arbitrary df's. *Statistica Neerlandica*, **35**, 169–189.

_________ (1982). Bootstrapping robust regression. *Comm. Statist. Theor. Meth.* **A11** (9), 1205–1224.

_________ (1991). Embedding the finite sampling process at a rate. *Ann. Probab.* **19**, 826–842.

Sen, P. K. (1966). On a distribution-free method of estimating asymptotic efficiency of a class of nonparametric tests. *Ann. Math. Statist.* **37**, 1759–1770.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics.* J. Wiley & Sons, Inc., N. Y.

Singh, K. (1981). On asymptotic accuracy of Efron's bootstrap. *Ann Statist.* **9**, 1187–1195.

Sinha, A. and Sen, P.K. (1979). Progressively censored tests for clinical experiments and life testing problems based on weighted empirical distributions. *Comm. Statist. Theor. and Meth.* **A8**, 871–898.

Smirnov, N.V. (1947). Sur un critēre de symétrie de la loi de distribution d'une variable aléatoire. *Compte Rendu* (*Doklady*) *de l'Academic des Sciences de l'URSS* **LVI**, NO. 1.

Stephens, M.A. (1976). Asymptotic results for goodness-of-fit statistics with unknown parameters. *Ann Statist.* **4**, 357–369.

_________ (1979). Tests of fit for the logistic distribution based on the empirical df. *Biometrica* **66**, 3, 591–595.

Vanderzanden, A.J. (1980). Some results for the weighted empirical processes concerning the laws of iterated logarithm and weak convergence. Ph.D. Thesis, MSU.

_________ (1984). A functional law of the iterated logarithm for the weighted empirical processes. *J. Ind. Statist. Assoc*, **22**, 97–110.

Van Eeden, C. (1972). An analogue, for signed rank statistics, of Jurečková's asymptotic linearity theorem for rank statistics. *Ann. Math. Statist.* **43**, 791–802.

Williamson, M. A. (1979). Weighted empirical–type estimators of regression parameter. Ph.D. Thesis, MSU.

_________ (1982). Cramer–von Mises estimations of regression parameter: The rank analogue. *J. Mult. Anal.* **12**, 248–255.

Withers, C.S. (1975). Convergence of empirical processes of mixing r.v.'s on [0, 1]. *Ann. Statist.* **3**, 1101–1108.

Wolfowitz, J. (1953). Estimation by minimum distance method. *Ann. Inst. Stat. Math.* **5**, 9–23.

_________ (1954). Estimation of the components of stochastic structures. *Proc. Nat. Acad. Sci.* **40**, 602–606.

_________ (1957). Minimum distance estimation method. *Ann. Math. Statist.*, **28**, 75–88.